

UNIVERSIDADE FEDERAL DO AMAZONAS - UFAM

INSTITUTO DE COMPUTAÇÃO- ICOMP

PROGRAMA PÓS-GRADUAÇÃO EM INFORMÁTICA - PPGI

Ensino de Engenharia de Dados nas Universidades Brasileiras e o Mercado: Estado Atual e Propostas de Modernização

Tarsis Figueredo Azevedo

Manaus - AM

Dezembro de 2024

Tarsis Figueredo Azevedo

Ensino de Engenharia de Dados nas Universidades Brasileiras e o Mercado: Estado Atual e Propostas de Modernização

Dissertação de Mestrado submetida à avaliação, para a obtenção do título de Mestre em Informática no Programa de Pós-Graduação em Informática, Instituto de Computação.

Orientador(a)

Altigran Soares da Silva, D.Sc.

Universidade Federal do Amazonas - UFAM

Instituto de Computação- IComp

Manaus - AM

Dezembro de 2024

Ficha Catalográfica

Elaborada automaticamente de acordo com os dados fornecidos pelo(a) autor(a).

F475e Figueredo Azevedo, Tarsis
Ensino de Engenharia de Dados nas Universidades Brasileiras e o Mercado: Estado Atual e Propostas de Modernização / Tarsis Figueredo Azevedo. - 2024.
337 f. : il., color. ; 31 cm.

Orientador(a): Altigran Soares da Silva.
Dissertação (mestrado) - Universidade Federal do Amazonas, Programa de Pós-Graduação em Informática, Manaus, 2024.

1. Educação em Engenharia de Dados. 2. Análise de Currículo. 3. Necessidades da Indústria. 4. Análise Comparativa. 5. Lacuna entre Indústria e Academia. I. Silva, Altigran Soares da. II. Universidade Federal do Amazonas. Programa de Pós-Graduação em Informática. III. Título

---

Ministério da Educação
Universidade Federal do Amazonas
Coordenação do Programa de Pós-Graduação em Informática

FOLHA DE APROVAÇÃO

"ENSINO DE ENGENHARIA DE DADOS NA UNIVERSIDADES BRASILEIRAS E O MERCADO: ESTADO ATUAL E PROPOSTAS DE MODERNIZAÇÃO"

TARSIS FIGUEREDO AZEVEDO

DISSERTAÇÃO DE MESTRADO DEFENDIDA E APROVADA PELA BANCA EXAMINADORA CONSTITUÍDA PELOS PROFESSORES:

Prof. Dr. Altigran Soares da Silva - PRESIDENTE

Prof. Dr. Joao Marcos Bastos Cavalcanti - MEMBRO EXTERNO

Profa. Dra. Mirella Moura Moro - MEMBRO EXTERNO

Prof. Dr. Leandro Silva Galvão de Carvalho - MEMBRO EXTERNO

MANAUS, 17 de dezembro de 2024.

Documento assinado eletronicamente por **Altigran Soares da Silva**, **Professor do Magistério Superior**, em 17/02/2025, às 08:17, conforme horário oficial de Manaus, com fundamento no art. 6º, § 1º, do Decreto nº 8.539, de 8 de outubro de 2015.

Documento assinado eletronicamente por **Mirella Moura Moro**, **Usuário Externo**, em 17/02/2025, às 13:41, conforme horário oficial de Manaus, com

fundamento no art. 6º, § 1º, do Decreto nº 8.539, de 8 de outubro de 2015.

Documento assinado eletronicamente por **Leandro Silva Galvão de Carvalho**, **Professor do Magistério Superior**, em 12/03/2025, às 12:00, conforme horário oficial de Manaus, com fundamento no art. 6º, § 1º, do Decreto nº 8.539, de 8 de outubro de 2015.

Documento assinado eletronicamente por **João Marcos Bastos Cavalcanti**, **Professor do Magistério Superior**, em 12/03/2025, às 14:41, conforme horário oficial de Manaus, com fundamento no art. 6º, § 1º, do Decreto nº 8.539, de 8 de outubro de 2015.

Documento assinado eletronicamente por **Maria do Perpétuo Socorro Vasconcelos Palheta**, **Secretária**, em 16/03/2025, às 19:38, conforme horário oficial de Manaus, com fundamento no art. 6º, § 1º, do Decreto nº 8.539, de 8 de outubro de 2015.

A autenticidade deste documento pode ser conferida no site https://sei.ufam.edu.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **2375434** e o código CRC **736DAFB6**.

Avenida General Rodrigo Octávio, 6200 - Bairro Coroado I Campus Universitário Senador Arthur Virgílio Filho, Setor Norte - Telefone: (92) 3305-1181 / Ramal 1193
CEP 69080-900, Manaus/AM, coordenadorppgi@icomp.ufam.edu.br

Dedico este trabalho ao meu pai, que nunca deixou de sonhar com um futuro melhor através da educação.

AGRADECIMENTOS

Gostaria de expressar minha profunda gratidão a Deus, cuja força e esperança foram meu alicerce em cada momento desafiador desta jornada. À minha esposa, pelo apoio incondicional, paciência e compreensão ao longo de todo o período, a ela meu amor e reconhecimento eterno. Ao meu orientador, agradeço pela paciência, pela parceria e pelos ensinamentos que tanto contribuíram para a realização deste trabalho. Cada um de vocês teve um papel indispensável na concretização desta conquista, e sou imensamente grato por terem estado ao meu lado.

Eu sei de muito pouco. Mas tenho a meu favor tudo o que não sei e – por ser um campo virgem – está livre de preconceitos. Tudo o que não sei é a minha parte maior e melhor: é a minha largueza. É com ela que eu compreenderia tudo. Tudo o que não sei é que constitui a minha verdade.

Clarice Lispector

Ensino de Engenharia de Dados nas Universidades Brasileiras e o Mercado: Estado Atual e Propostas de Modernização

Autor: Tarsis Figueredo Azevedo

Orientador: Altigran Soares da Silva, D.Sc.

Resumo

O termo Engenharia de Dados (ED) tem sido utilizado com frequência na literatura para se referir aos processos de adquirir, organizar e preparar dados para serem consumidos em análises exploratórias, como entrada de sistemas e aplicações ou outros contextos similares. Com o aumento do volume de dados, atualmente, e sua projeção de crescimento, as tecnologias de ED tiveram avanços nas últimas décadas e continuam evoluindo em uma velocidade alta. Ao mesmo tempo, existe a percepção de que tais avanços tecnológicos ainda não foram absorvidos pela academia, que ainda mantém um foco em Gerenciamento de Dados Relacionais na maior parte das matrizes curriculares. Portanto, nesse trabalho de mestrado, definimos uma abordagem para investigar e mitigar essa defasagem, que consiste em duas etapas. Na primeira, realizamos extenso levantamento sobre as disciplinas, bibliografias e ementas relacionadas à ED, em 23 universidades brasileiras e junto a empresas de tecnologia do país. Esses levantamentos revelaram que, de fato, existe uma lacuna entre o ensino de ED e as demandas do mercado, com currículos acadêmicos frequentemente desatualizados quanto a tópicos considerados relevantes para a indústria contemporânea. Na segunda etapa do trabalho, propomos uma alternativa moderna para o ensino de ED em universidades, tanto do ponto de vista do conteúdo quanto da forma de ensino, inspirados por experiências modernas de ensino de tópicos avançados de ED.

Palavras-chave: Educação em Engenharia de Dados, Análise de Currículo, Necessidades da Indústria, Análise Comparativa, Metodologias Educacionais, Lacuna entre Indústria e Academia.

Lista de ilustrações

Figura 1 – Amostra da taxonomia de tópicos da bibliografia. 34
Figura 2 – Amostra da taxonomia de tópicos das ementas. 35
Figura 3 – Amostra da taxonomia de tópicos da indústria. 39
Figura 4 – Distribuição dos Alunos da Turma por Faixa Etária 58
Figura 5 – Distribuição dos alunos que trabalham na área de computação. 59
Figura 6 – Avaliação do Conteúdo Aplicado 60
Figura 7 – Avaliação da Disciplina 62
Figura 8 – Avaliação da relevância e aplicabilidade 63

Lista de tabelas

Tabela 1 – Levantamento das habilidades esperadas dos engenheiros de dados. 37
Tabela 2 – Ementa da nova disciplina de engenharia de dados 46
Tabela 3 – Unidades programadas para nova disciplina. 47
Tabela 8 – Tópicos de Bancos de Dados e Computação 296
Tabela 9 – Tópicos e Frequências dos Capítulos 306
Tabela 10 – Pesquisa junto a industria brasileira de tecnologia 311
Tabela 11 – Tópicos de Bancos de Dados e Computação 329
Tabela 12 – Pesquisa Qualitativa da Disciplina de Engenharia de Dados ENGD 2024-01 . 335
Tabela 13 – Avaliação da Disciplina de Engenharia de Dados 337

SUMÁRIO

1 INTRODUÇÃO 11

2 METODOLOGIA 18
2.1 Definição do conceito de Engenharia de Dados (ED) 19
2.2 Levantamento das disciplinas de ED no ensino superior 19
2.3 Normalização das informações coletadas 20
2.4 Construção de taxonomias de tópicos acadêmicos 20
2.5 Levantamento com líderes de empresas de tecnologia 21
2.6 Comparação entre ensino acadêmico e demandas do mercado . . . 22
2.7 Criação de uma disciplina de Engenharia de Dados 23
2.8 Aplicação da nova disciplina e avaliação dos resultados 23

3 TRABALHOS RELACIONADOS 25

4 CRIAÇÃO DA TAXONOMIA DA ACADEMIA E DA INDÚSTRIA 33
4.1 Levantamento das Bibliografias e Ementas das Disciplinas 33
4.2 Análise das necessidades da indústria brasileira em relação à Engenharia de Dados 36
4.3 Discussão 38

5 ENGENHARIA DE DADOS – UMA NOVA DISCIPLINA 42
5.1 Objetivo Geral 43
5.2 Objetivos específicos 44
5.3 Definição da Ementa 44
5.4 Detalhamento da ementa 47

6 APLICAÇÃO DA NOVA DISCIPLINA 55
6.1 Planejamento das Aulas e Aplicação no Semestre 55
6.2 Avaliações Teóricas e Trabalhos Práticos 56

6.3 Avaliação dos alunos . . . 57
6.3.1 Avaliação do Conteúdo . . . 57
6.3.2 Discussão dos resultados . . . 63
6.4 Avaliação da Indústria . . . 65
6.4.1 Discussão dos resultados . . . 66

7 CONCLUSÃO E TRABALHOS FUTUROS . . . 68

Referências . . . 72

APÊNDICE A LISTA COMPLETA DAS DISCIPLINAS COM EMENTA E BIBLIOGRAFIA . . . 75

APÊNDICE B LISTA DE DISCIPLINAS SELECIONADAS . . . 116

APÊNDICE C BIBLIOGRAFIA DAS DISCIPLINAS COMPLETA E NORMALIZADA . . . 125

APÊNDICE D BIBLIOGRAFIA DEDUPLICADA . . . 183

APÊNDICE E CAPÍTULOS SELECIONADOS DA BIBLIOGRAFIA LEVANTADA . . . 202

APÊNDICE F TAXONOMIA DA BIBLIOGRAFIA . . . 296

APÊNDICE G TAXONOMIA DAS EMENTAS . . . 306

APÊNDICE H DADOS BRUTOS DA PESQUISA HABILIDADES PARA ENGENHEIROS DE DADOS . . . 311

APÊNDICE I TAXONOMIA DA INDÚSTRIA . . . 329

APÊNDICE J BIBLIOGRAFIA NOVA COMPLETA . . . 332

APÊNDICE K PESQUISA QUALITATIVA DA DISCIPLINA DE ENGENHARIA DE DADOS . . . 335

APÊNDICE L AVALIAÇÃO DA DISCIPLINA DE ENGENHARIA DE DADOS . . . 337

1

INTRODUÇÃO

A digitalização acelerada de praticamente todas as atividades da sociedade moderna teve como uma de suas maiores consequências a produção de gigantescos repositórios de dados, em um volume sem precedentes na história. É tamanha a prevalência dos dados que eles são atualmente comparados em importância de commodities como o petróleo, ou mesmo a bens de capital (ARRIETA-IBARRA et al., 2018). De fato, a quantidade de dados gerados globalmente aumenta diariamente, com estimativa de crescimento de 459 bilhões de dólares entre 2024 e 2028, com taxa de crescimento anual composta de 55%, impulsionado pela adoção em massa de dispositivos IoT, redes sociais e outras tecnologias digitais, possibilitando negócios de ganhar importantes percepções e aprendizados a partir do consumo desses dados. Entretanto, desafios como segurança, integrações complexas de dados e o alto custo de implementação continuam a existir.(TECHNAVIO, 2024)

Além disso, novas arquiteturas e tecnologias de dados estão surgindo para lidar com esse volume, como é o caso dos Data Lakehouses (ZAHARIA et al., 2021) e de plataformas como Apache Iceberg[1]. Ademais, a sociedade em geral e o governo podem se beneficiar do uso massivo de dados em várias áreas, como a da segurança pública, a das aplicações em saúde e a do desenvolvimento sustentável, auxiliando a criação de novas políticas, a distribuição de recursos, as cidades inteligentes e a aplicação de medidas preventivas, levando a maior eficiência no uso da máquina pública, e trazendo transparência para a administração de recursos (LEE, 2020).

Assim, um importante desafio do nosso tempo é a disponibilidade de mão de obra capacitada para lidar com esse volume de dados, e, mais ainda, poder desenvolver soluções

[1] https://iceberg.apache.org/

para processá-los de forma efetiva e escalável. para gerar, a partir desses dados, conhecimento que possa ser usado em benefício das pessoas, das Instituições, das empresas e da sociedade. Em particular, é bem conhecido que 80% do trabalho envolvido nesse tipo de atividade está associado a tarefas como integrar, preparar, transformar e manipular em grande e pequena escala os dados a serem utilizados (BIE et al., 2022). Por essa razão, neste trabalho, empregamos o termo de Engenharia de Dados (ED) para nos referirmos a essas atividades. Assim, nosso ponto de vista é que, para atender adequadamente às demandas da sociedade e do mercado, é muito importante que os profissionais de computação, formados pelas universidades, dominem os principais métodos e técnicas de Engenharia de Dados.

Objetivos

Este trabalho tem como objetivo geral investigar a adequação entre o ensino de Engenharia de Dados (ED) nas universidades brasileiras e as demandas atuais da sociedade e da indústria, propondo recomendações e implementações para alinhar os currículos acadêmicos da área às necessidades emergentes do mercado e sociedade. Além disso, o trabalho tem também os seguintes objetivos específicos:

- Mapear o estado atual do ensino de Engenharia de Dados nas universidades brasileiras: levantar as disciplinas da área de gerenciamento de dados disponibilizadas pelos cursos de Ciência da Computação, de Engenharia de Sistemas, de Sistemas da Informação e afins, coletando suas ementas e bibliografias.

- Identificar os tópicos atualmente abordados em disciplinas relacionadas à ED. Analisar a bibliografia utilizada e a frequência de tópicos ensinados. Identificar as demandas da indústria em relação à formação em Engenharia de Dados: entender junto a lideranças de empresas de tecnologia quais as habilidades essenciais para um Engenheiro de Dados recém-formado.

- Levantar as habilidades técnicas consideradas essenciais pelas empresas de tecnologia no mercado e na sociedade brasileira. Construir uma taxonomia das competências demandadas para atuação profissional na área. Comparar o conteúdo acadêmico com as demandas do mercado e da sociedade: construir taxonomias dos tópicos mais relevantes para indústria e academia, identificando as discrepâncias entre eles.

- Identificar discrepâncias entre o que é ensinado nas universidades e as necessidades do atuais da indústria e da sociedade. Determinar lacunas específicas na formação de engenheiros de dados recém-graduados. Propor melhorias no currículo acadêmico: atualizar o conteúdo e a forma como o gerenciamento de dados é ensinado na academia, trazendo as inovações da indústria para dentro da academia. Adicionando a prática deliberada para engajar os alunos.

- Desenvolver uma disciplina experimental de Engenharia de Dados baseada nos tópicos identificados como prioritários. Elaborar uma ementa que equilibre teoria e prática, com ênfase em métodos modernos e aplicações reais. Testar e avaliar a viabilidade da proposta curricular: aplicar a proposta da disciplina para alunos de Ciência da Computação de uma universidade a fim de entender a viabilidade desse modelo.

- Implementar a disciplina experimental em cursos de graduação em Ciência da Computação e Engenharia de Software. Coletar feedback de alunos, profissionais da indústria e sociedade sobre a eficácia da nova abordagem. Propor inovações no método de ensino: desenvolver novos materiais didáticos, trazendo exemplos práticos da aplicação das técnicas e tecnologias utilizadas no mercado para ilustrar aos alunos sua aplicabilidade. Desenvolver projetos para que os alunos possam praticar, resolvendo problemas reais, com o que aprenderão nas aulas.

- Incorporar atividades práticas, como laboratórios práticos, miniprojetos e um projeto final aplicado. Avaliar o impacto dessas metodologias no engajamento e na absorção do conteúdo pelos alunos. Divulgar os resultados obtidos: criar materiais práticos para que os alunos possam praticar o que estão aprendendo. Aplicar os projetos desenvolvidos e avaliar a implementação dos alunos. Aplicar um questionário ao final do semestre para coletar os *feedbacks* dos alunos. Consolidar os resultados no trabalho.

- Publicar os achados da pesquisa em periódicos e eventos científicos relevantes para a área de Computação. Apresentar recomendações para reformulação curricular em conferências educacionais e técnicas.

Hipóteses

A adequação entre o ensino acadêmico e as demandas do mercado e da sociedade é um fator crucial para a formação de profissionais competentes na área de Engenharia de Dados, especialmente diante da crescente importância dessa disciplina para a sociedade e a economia. Apesar da relevância do tema, evidências sugerem que os avanços industriais em Engenharia de Dados não têm sido acompanhados de maneira satisfatória pelos currículos das universidades brasileiras. Este desalinhamento compromete a capacidade dos recém-graduados de atender às exigências técnicas do mercado.

Nesse contexto, e alinhados aos objetivos deste trabalho, formulamos hipóteses para investigar a existência de uma possível discrepância entre o que é ensinado nas universidades e as necessidades da indústria, além de explorar caminhos viáveis para a atualização curricular e a melhoria dos métodos de ensino. As hipóteses a seguir direcionam a pesquisa, permitindo uma análise sistemática das lacunas e a proposição de soluções fundamentadas e práticas.

- **Hipótese 1:** Existe uma discrepância entre o conteúdo ensinado nas universidades brasileiras com relação à engenharia dados e as necessidades atuais da indústria
  - **Pergunta 1.1:** O que é ensinado atualmente nas principais universidades brasileiras nessa área?
  - **Pergunta 1.2:** Quais as principais necessidades atuais da indústria nessa área?
- **Hipótese 2:** Considerando que tal discrepância existe, é possível atualizar o currículo e o método de ensino atuais para torná-los mais adequados às necessidades da indústria
  - **Pergunta 2.1:** Nesse sentido, o que é necessário atualizar nos currículos em vigor?
  - **Pergunta 2.2:** É possível ministrar o conteúdo necessário em uma disciplina de curso de graduação em Ciência da Computação ou similares?
  - **Pergunta 2.3:** Alguma outra mudança teria de ser feita nos currículos para absorver essa atualização?
  - **Pergunta 2.4:** É possível utilizar novos métodos de ensino para facilitar a aprendizagem do conteúdo e melhorar o engajamento dos alunos?

Desenvolvimento do Trabalho

Para investigar a Hipótese 1, procuramos compreender o que é atualmente ensinado sobre Engenharia de Dados nas principais universidades brasileiras. Para isso, foi necessário realizar um levantamento dos cursos ofertados na área e da bibliografia utilizada em disciplinas relacionadas. Com esses dados, buscou-se obter uma visão geral do conteúdo transmitido aos alunos de graduação sobre esse tópico. Além disso, procuramos identificar as necessidades da indústria, por meio de uma pesquisa junto a empresas brasileiras de tecnologia da informação, que trabalham com dados massivos. O objetivo foi elencar as habilidades demandadas para que recém-graduados possam atuar nessa área. A partir dessa análise, pretendeu-se avaliar se há uma discrepância entre o conteúdo ensinado nas universidades brasileiras, em relação à Engenharia de Dados e às demandas atuais do mercado.

Para investigar a Hipótese 2, exploramos e discutimos possíveis mudanças nos currículos dos cursos de computação. O foco esteve em analisar a viabilidade de criar uma nova disciplina que atendesse às demandas identificadas na pesquisa. Essa disciplina seria estruturada para incluir material didático atualizado, tópicos relevantes para Engenharia de Dados com conteúdo às necessidades emergentes da área, mantendo-se alinhada à carga horária disponível nos currículos atuais.

Adicionalmente, propusemos a adoção de métodos de ensino mais interativos, como atividades práticas assistidas em laboratório; miniprojetos ao longo do semestre; e um projeto final que avalie a aplicação prática dos conhecimentos adquiridos. Essas abordagens seriam avaliadas com base nos resultados obtidos na pesquisa, garantindo alinhamento às demandas do mercado sem comprometer os objetivos acadêmicos.

Resultados do Trabalho

Como resultado da investigação da primeira hipótese e suas perguntas, realizamos uma pesquisa para determinar quais assuntos sobre Engenharia de Dados (ED) são atualmente ensinados nas universidades brasileiras. O nosso estudo se concentrou nos cursos de graduação na área de computação das universidades brasileiras, em que tipicamente os tópicos de ED são cobertos em disciplinas intituladas "Bancos de Dados"ou variações desse nome. Para determinar tais assuntos, realizamos um recorte, considerando as universidades com grupos de pesquisa mais

tradicionais em ED, analisando as ementas dessas disciplinas, o que resultou em uma lista. Com essa lista, fizemos um levantamento da bibliografia utilizada. E, a partir dessas informações, construímos uma taxonomia de tópicos atualmente ensinados nessas disciplinas.

Com apoio na taxonomia construída, baseada na bibliografia, elaboramos uma taxonomia de tópicos extraídos das ementas, identificando os assuntos destacados em cada uma delas. Ainda foi feita uma análise de frequência da ocorrência de cada tópico nas ementas para compreender a sua prevalência. Vale ressaltar que, para essa análise, os nomes dos tópicos foram normalizados.

Ao mesmo tempo, para determinar como esses assuntos abrangem as necessidades da sociedade e do mercado brasileiros, realizamos um levantamento com várias das principais empresas do setor "tecnologia da informação", buscando identificar habilidades técnicas necessárias para que um engenheiro de dados recém-graduado atenda às demandas do mercado. Com essas informações, desenvolvemos uma nova taxonomia de tópicos, que representam as habilidades essenciais para a indústria brasileira de tecnologia na área de dados.

Com essas duas taxonomias em mãos, realizamos uma comparação sistemática entre elas para identificar as diferenças entre os tópicos já cobertos pelas universidades e aqueles que ainda faltam para atender às demandas da indústria. Destaca-se que, conforme o levantamento bibliográfico que fizemos e que detalhamos no Capítulo 4, há poucos estudos similares no mundo — mencionados na Seção 3 — e nenhum realizado no Brasil até o momento. Assim, esta pesquisa se mostra relevante por oferecer uma visão das lacunas no ensino e na formação dos profissionais em Engenharia de Dados no país.

A partir dos resultados obtidos, foi proposta uma nova disciplina que chamamos "Engenharia de Dados", com foco nas habilidades identificadas como cruciais para a indústria. De forma experimental, essa disciplina foi oferecida para alunos do 6º período dos cursos de Ciência da Computação e de Engenharia de Software, da Universidade Federal do Amazonas (UFAM), usando, no entanto, as denominações de disciplinas já existentes nos currículos desses cursos, respectivamente, "Bancos de Dados II"e "Prática em Bancos de Dados".

A disciplina utilizou uma bibliografia atualizada, incorporando os conceitos mais modernos de Engenharia de Dados. Além disso, o fluxo das aulas foi estruturado para adotar ciclos curtos que alternavam aula teórica, atividades práticas em laboratório e avaliações, com o objetivo de manter os alunos engajados. O curso culminou em um projeto aplicado, baseado em cenários reais, no qual os alunos puderam integrar e aplicar os tópicos ensinados ao longo

do semestre. Ao final da disciplina, foi realizada uma pesquisa com os alunos para coletar suas opiniões sobre a nova disciplina. Além disso, conduzimos uma pesquisa junto à indústria e à sociedade para verificar se a nova disciplina atende às demandas e necessidades atuais do mercado.

Como resultados deste trabalho, dois artigos foram desenvolvidos e publicados, apresentando análises parciais da investigação realizada. O primeiro artigo, publicado nos anais do Simpósio Brasileiro de Bancos de Dados (SBBD) 2023 (AZEVEDO; SILVA, 2023), explora o ensino de Engenharia de Dados (ED) nas universidades brasileiras, analisando disciplinas e bibliografias à luz das demandas do mercado, com destaque para lacunas em temas como "gerência de dados em nuvem"e "*workflows* de dados". O segundo artigo, publicado no Simpósio Brasileiro de Educação em Computação (EduComp) 2024 (AZEVEDO; SILVA, 2024), expande o escopo do levantamento para 23 universidades, consolidando as conclusões do primeiro estudo e propondo recomendações para a reformulação curricular, com foco na integração de ED no contexto da educação em computação.

Organização do trabalho

O restante desta dissertação está organizada da seguinte forma: no Capítulo 2, apresentamos detalhes sobre o levantamento feito nas universidades brasileiras, mostramos como foi realizado um levantamento com as empresas de tecnologia e, como, a partir disso, criamos as taxonomias de tópicos atuais e necessários. Também é apresentado como as aulas foram desenvolvidas e como a pesquisa final com alunos e a indústria foi conduzida. No Capítulo 3, apresentamos um levantamento de trabalhos relacionados à nossa pesquisa, suas contribuições e em que nossa pesquisa difere deles.

No Capítulo 4, detalhamos os resultados obtidos para validar a primeira hipótese e suas perguntas. No Capítulo 5, descrevemos a nova disciplina de Engenharia de Dados, e detalhamos seus objetivos e ementa. Depois, no Capítulo 6, apresentamos como a disciplina foi aplicada, bem como suas avaliações e trabalho final, e por fim, apresentamos o resultado da avaliação de alunos, que foram pessoas da indústria e da sociedade civil. E, finalmente, temos o Capítulo 7, em que apresentamos as nossas conclusões deste trabalho e as possibilidades de trabalhos futuros.

2

METODOLOGIA

Neste capítulo, apresentamos a metodologia adotada nesta pesquisa. As etapas do estudo foram organizadas em um conjunto de fases estruturadas, descritas a seguir:

1. **Definição do conceito de Engenharia de Dados (ED)**: Estabelecimento de uma base conceitual para o estudo.
2. **Levantamento das disciplinas de ED no ensino superior**: Coleta de bibliografias e ementas de disciplinas relacionadas em instituições brasileiras.
3. **Normalização das informações coletadas**: Organização e padronização das bibliografias e ementas para análise comparativa.
4. **Construção de taxonomias de tópicos acadêmicos**: Desenvolvimento de taxonomias com base nas informações normalizadas.
5. **Levantamento com líderes de empresas de tecnologia**: Identificação das habilidades e tópicos valorizados na contratação de engenheiros de dados.
6. **Comparação entre ensino acadêmico e demandas do mercado**: Análise das lacunas e divergências entre as taxonomias acadêmicas e as exigências profissionais.
7. **Criação de uma disciplina de Engenharia de Dados**: Proposta de um nova disciplina, com base nos resultados anteriores da pesquisa.
8. **Aplicação da nova disciplina e avaliação dos resultados**: Aplicação da proposta de nova disciplina e avaliação dos resultados dessa aplicação na visão dos alunos e da indústria.

A descrição detalhada de cada uma dessas etapas será apresentada ao longo deste capítulo.

2.1 Definição do conceito de Engenharia de Dados (ED)

Neste trabalho definimos "Engenharia de Dados"(ED) como o processo de adquirir, organizar e preparar dados para serem consumidos em análises exploratórias, como entrada de algoritmos ou outros contextos (BIE et al., 2022). Em um nível mais abstrato, a ED resolve problemas relacionados à organização e qualidade dos dados, bem como à extração de características de entidades do mundo real representadas em meio digital (NAZÁBAL et al., 2020).

2.2 Levantamento das disciplinas de ED no ensino superior

Com base na definição de ED, selecionamos disciplinas que se relacionam diretamente com esses problemas. Tipicamente, estas disciplinas são aquelas chamadas de “Bancos de Dados 1”, “Bancos de Dados 2”, “Mineração de Dados”, “Tópicos Especiais em Bancos de Dados”, “Computação em Nuvem”, entre outras. Vale ressaltar que as disciplinas que não abordam diretamente esses problemas, como “Inteligência Artificial” e “Ciência de Dados”, foram excluídas de nossa lista. A lista completa das disciplinas selecionadas pode ser encontrada no Apêndice B.

Conseguimos fazer o levantamento de 108 disciplinas, e dessas, listamos suas ementas e suas bibliografias. Descartamos as disciplinas que consideramos não relacionadas a ED com base na nossa definição de engenharia de dados, ficando a lista final com 90 disciplinas. Nesse levantamento, algumas universidades não puderam ser incluídas devido a diferentes motivos, tais como falta de informações disponíveis no site, indisponibilidade de currículo, links inativos no site e currículos sem ementa ou bibliografia. Além disso, outros desafios encontrados no processo de levantamento foram: sites confusos e de difícil navegação; além de ementas incompletas, muito vagas ou indisponíveis; e bibliografias não padronizadas ou indisponíveis. Sabemos que tais problemas podem impactar na confiabilidade da pesquisa, mas acreditamos ter gerado dados suficientes para seguir com a análise. A lista completa de disciplinas com suas ementas e bibliografias pode ser vista no Apêndice A.

2.3 Normalização das informações coletadas

Para normalizar os itens das bibliografias, decidimos usar o seu ISBN. Porém, nem todas as bibliografias disponibilizadas pelas universidades continham esse identificador. Para levantar os faltantes, fizemos uma busca no site da Agência Brasileira do ISBN, que faz parte da Câmara Brasileira do Livro (CBL), porém, nem todos os livros estavam disponíveis nesse site: alguns, por serem edições internacionais, e outros por algum outro motivo. Por isso, usamos o site ISBN Search (isbnsearch.org) para completar as buscas. Nessas buscas, nem sempre foi possível achar a edição exata do livro, e por isso, decidimos usar o ISBN da edição mais próxima à desejada.

Também, nem sempre foi possível encontrar a edição brasileira do livro, como listada na bibliografia, e por isso, optamos por utilizar o ISBN da versão original, e colocamos essa informação na coluna "Edição Brasileira", no arquivo disponibilizado. Em algumas disciplinas, encontramos na bibliografia materiais de consulta, periódicos, revistas, congressos e notas de aula. Para esses, não colocamos dados de ISBN, autor e título, e descartamos das análises para taxonomia. Esse levantamento completo pode ser visto no Apêndice C.

2.4 Construção de taxonomias de tópicos acadêmicos

Para construir uma taxonomia dos tópicos ensinados nas disciplinas de Engenharia de Dados, nas universidades brasileiras, com base em suas bibliografias, primeiro fizemos a deduplicação da lista de livros, resultando numa lista final que pode ser vista no Apêndice D, depois fizemos um levantamento dos capítulos de cada um dos livros, totalizando 874 capítulos, e investigamos os assuntos abordados em cada um deles, por meio de seus sumários. Organizamos essas informações no Apêndice E, que serviu como base para a criação da taxonomia de tópicos da bibliografia. Tal taxonomia tem 3 níveis, de tópicos mais genéricos a mais específicos, e pode ser vista completamente no Apêndice F. Ressalta-se que a intenção deste trabalho não foi criar uma taxonomia definitiva para tópicos relacionados a "dados", mas sim, gerar uma visão geral do que é ensinado nas universidades brasileiras na área. Embora não tenhamos encontrado o sumário de 10 desses livros (7%), consideramos que isso não prejudicou a representatividade do trabalho, uma vez que acreditamos que os assuntos tratados por esses já estão incluídos em outros livros que encontramos.

A partir dessa taxonomia, podemos construir outra taxonomia de tópicos das ementas

das disciplinas ligadas à Engenharia de Dados nas universidades brasileiras, com base nas informações disponibilizadas, em que cada assunto foi mapeado para um tópico da taxonomia da bibliografia, descrita acima. Isso foi feito porque entendemos que os assuntos tratados nas ementas são um subconjunto de todos os assuntos tratados na bibliografia de cada disciplina. Ressalta-se que essa taxonomia não tem intenção de afirmar o que é exatamente ensinado em cada uma dessas disciplinas, dado que a ementa é só uma representação simplista do conteúdo dado em sala de aula. Embora não tenhamos ementas em 16 das disciplinas (9%), consideramos que isso não prejudicou a representatividade da nossa análise.

Após a construção da taxonomia de tópicos com base nas ementas disponibilizadas, realizamos uma análise de frequência nas ementas para compreender como os tópicos listados foram utilizados pelas disciplinas. Para isso, procuramos os tópicos listados na taxonomia em cada ementa, e para cada ocorrência, somamos 1 à frequência. Alguns termos foram agrupados sob um mesmo tópico, como os tópicos "Normalização", "Formas Normais", "1FN", "2FN" e "3FN", que estão todos sob o tópico "Formas Normais". Embora reconheçamos que essa análise é superficial e limitada, consideramos ser relevante para este estudo, uma vez que fornece indícios do que é ensinado nas disciplinas. Essa taxonomia pode ser vista completa no Apêndice G.

Com essa taxonomia, foi possível obter uma lista dos tópicos mais frequentes tratados nas disciplinas de ED das universidades pesquisadas. É importante destacar que algumas disciplinas têm ementas muito abertas, especialmente as disciplinas como "Tópicos Avançados" ou "Tópicos Especiais" em bancos de dados. Essas disciplinas podem ser usadas para tratar de temas relevantes e atuais, embora não possamos afirmar isso com certeza nesta pesquisa, pois não tivemos acesso ao material dado em sala pelos professores de cada disciplina. Além disso, ressaltamos que essa taxonomia não representa exatamente os assuntos tratados em sala. Para uma pesquisa que reflita exatamente o que é ensinado na sala de aula, precisamos de acesso ao Plano de Ensino e ao Plano de Aula de cada uma.

2.5 Levantamento com líderes de empresas de tecnologia

Para determinar como esses assuntos abrangem as necessidades da sociedade e do mercado brasileiro, foi realizado também um levantamento com várias das principais empresas do ramo de tecnologia da informação do país. Nesse levantamento, fizemos 3 perguntas para mapear essas necessidades. A primeira pergunta foi feita em forma de múltipla escolha para determinar

o nível de habilidade desejável para um Engenheiro de Dados recém-formados, de uma lista de tópicos pré-selecionados por nós. Assim, em nível teórico, pretendeu-se uma exposição discursiva a um assunto prático-básico, ou seja, a utilização de tal habilidade em um problema real simples; e, em nível prático avançado, a utilização de tal habilidade em problemas reais complexos. Também fizemos uma pergunta discursiva para ser possível justificar as escolhas na primeira pergunta; e uma terceira pergunta, também discursiva, para ser possível listar habilidades não listadas por nós na primeira pergunta. Tal levantamento foi distribuído por e-mail para pessoas escolhidas por nós, em posições de liderança dentro dessas empresas. E seu resultado bruto pode ser visto no Apêndice H.

A partir dos resultados desse levantamento, geramos uma taxonomia de 79 tópicos, em 3 níveis, do mais genérico ao mais específico, considerados relevantes para a indústria na contratação de engenheiros de dados, que pode ser conferida no Apêndice I. Para completar essa taxonomia, utilizamos outras quatro fontes. A primeira fonte foi o modelo de conceitos-chave para gerenciamento de dados criado por (GRILLENBERGER; ROMEIKE, 2017), que traz quatro categorias principais: Práticas, Princípios de Projeto, Mecanismos e Tecnologias Centrais, com termos genéricos e bastante abrangentes, porém, que definem o estado da Engenharia de Dados atual. A taxonomia também foi complementada pelos tópicos propostos em (BIE et al., 2022) nos quais são identificadas necessidades do dia a dia de ED.

Outra fonte consultada para a criação da taxonomia foi o currículo de computação da ACM, em 2023 (ACM; AAAI, 2024). Por fim, também foi usado o conhecimento do primeiro autor desta dissertação na indústria, como engenheiro de dados, com mais de 12 anos de experiência em grandes empresas, e também sua experiência tendo fundado e ensinado em um *bootcamp* de engenharia e análise de dados durante 6 anos, sendo a criação desse *bootcamp* motivada pela falta de profissionais capacitados na área de ED no mercado, em 2017.

2.6 Comparação entre ensino acadêmico e demandas do mercado

Nesta etapa, realizamos uma análise comparativa entre a taxonomia de ementas, com sua frequência, e a taxonomia de tópicos da indústria. Identificamos e destacamos as diferenças entre as duas taxonomias, visando aperfeiçoar a nossa compreensão sobre os tópicos mais

relevantes na área de Engenharia de Dados e identificar diferenças relevantes entre o que é ensinado e as necessidades da indústria.

2.7 Criação de uma disciplina de Engenharia de Dados

Uma vez identificadas as lacunas da academia, apresentamos a proposta de uma nova disciplina a ser ministrada para cursos de graduação em ciência da computação como uma estratégia para mitigar as deficiências identificadas na primeira parte deste trabalho. Essa proposta utilizou como base os resultados preliminares detalhados na Seção 4.3.

2.8 Aplicação da nova disciplina e avaliação dos resultados

Aplicamos essa nova disciplina para alunos do 6º período dos cursos de Ciência da Computação e Engenharia de Software da UFAM, substituindo as disciplinas de *Bancos de Dados II* e *Prática em Bancos de Dados*, atualizando os assuntos e revisando seu material. Dividimos os assuntos em unidades e preparamos materiais expositivos para apresentar os assuntos aos alunos. Dentro de cada aula, fizemos demonstrações práticas daquele assunto, seja mostrando uma ferramenta, um código e também uma plataforma de Nuvem Pública. Foram adicionados exemplos de aplicação de cada assunto no mundo real, visando tangibilizar o conhecimento.

Após um conjunto de unidades, fizemos um *Trabalho Prático*, para os alunos poderem aplicar os conhecimentos dados em aula. Também foram aplicadas duas avaliações teóricas para testar a absorção dos assuntos expostos. No final do curso, um projeto final prático foi pedido aos alunos, e, utilizando as soluções criadas nos trabalhos anteriores, eles deveriam criar uma plataforma de dados com processamento de ponta a ponta. A nova disciplina será descrita por completo no Capítulo 5.

Para avaliar a eficácia dessas mudanças na disciplina, foi feita uma pesquisa qualitativa com os alunos do semestre em que a análise foi desenvolvida. Tal pesquisa consistiu em uma série de perguntas e afirmações, com respostas na *escala Likert*, divididas em três seções. A primeira seção agrupou as perguntas e afirmações para avaliação da disciplina no quesito técnico, com o objetivo de avaliar o conteúdo de Engenharia de Dados da disciplina, de modo

geral. A segunda seção agrupou as perguntas e afirmações para avaliação da disciplina no quesito didático, com o objetivo de avaliar como o conteúdo foi dado durante o semestre. Na terceira seção, estavam as perguntas e afirmações exclusivas para alunos que já trabalham na área, para avaliar a aplicabilidade do conteúdo dado. Em cada seção, também foram feitas perguntas discursivas, para os alunos explicitarem livremente suas opiniões sobre o conteúdo, a disciplina e a aplicabilidade dos assuntos tratados no dia a dia. Os resultados dessa pesquisa foram apresentados na seção 6.3. A pesquisa foi distribuída no final do semestre para os alunos matriculados na disciplina.

Além da pesquisa com os alunos, foi feita uma pesquisa com a indústria e a sociedade. Tal pesquisa teve o objetivo de apresentar a disciplina e conseguir *feedbacks* valiosos de pessoas diretamente envolvidas com engenharia de dados. Foram feitas três perguntas, com respostas na escala Likert, cobrindo a relevância dos assuntos tratados na disciplina para o contexto atual da sociedade e da indústria, a aplicabilidade das habilidades no dia a dia das empresas e a entrada dos alunos no mercado de trabalho. Além dessas perguntas, tivemos duas de texto livre, uma para os respondentes listarem assuntos que deveriam ser tratados, mas não estavam na ementa, e outra para comentários gerais pertinentes. Os resultados dessa pesquisa foram apresentados na seção 6.4. Essa pesquisa foi distribuída por email.

Tais resultados são apresentados no Capítulo 6.

3

TRABALHOS RELACIONADOS

Neste capítulo, apresentamos alguns dos estudos mais relevantes na área de engenharia de dados quanto à definição dos seus conceitos-chave e características, e também sobre sua integração em cursos de Ciência da Computação, e ainda sobre formas de integrar metodologias de ensinos dinâmicas, como *bootcamps*, nas aulas de graduação. Avaliamos suas principais contribuições e realizamos comparações com a nossa própria pesquisa.

Segundo o "Computer Science Curricula 2023" da ACM, que estabelece diretrizes para currículos de Ciência da Computação (ACM; AAAI, 2024), desde a década de 1970, a área de gerenciamento de dados tem se concentrado principalmente em estudos de bancos de dados relacionais, com foco em tópicos como modelagem de dados, construção de consultas, processamento de consultas e estrutura interna dos sistemas de gerenciamento de bancos de dados (SGDBs). Hoje é esperado dos graduados em CC que eles tenham mais habilidades de usuários, e não de administradores de SGBDs. Podemos observar isso, quando analisamos trabalhos que visam catalogar conceitos-chave da computação, como "Great Principles of Computing" de Denning (DENNING, 2003), no qual o campo de gerenciamento de dados está restrito à "Recollection", tendo na sua descrição os importantes aspectos de Hierarquias, Persistência ou Compartilhamento, porém, muitos outros aspectos de gerenciamento de dados parecem sub-representados(GRILLENBERGER; ROMEIKE, 2017). Por outro lado, o volume de dados disponíveis e a ser gerenciado tem crescido exponencialmente a cada ano, aumentando a necessidade de profissionaisqualificados para analisar e lidar com esses dados. Nesse contexto, o mercado busca, cada vez mais, esses profissionais, enquanto as universidades se esforçam para formá-los (HASSAN; LIU, 2019).

Em um painel realizado durante a edição de 2021, da importante conferência *Very*

Large Databases (VLDB) sobre o futuro da educação em bancos de dados (IVES et al., 2021), os painelistas abordaram a evolução do ensino de bancos de dados diante do crescimento da área de ciência de dados. Eles destacaram a necessidade de adaptações curriculares para atrair e ensinar efetivamente à nova geração de estudantes, explorando as diferenças e intersecções entre Ciência da Computação (CC) e Ciência de Dados (CD). Foram debatidos o conteúdo e o foco dos cursos de bancos de dados e ciência de dados, considerando as mudanças nas exigências das áreas e os impactos nos estudantes e na comunidade acadêmica. Esse painel contribuiu significativamente para a discussão sobre a direção futura da educação em bancos de dados e sua integração, ou não, com a ciência de dados, fornecendo ideias e percepções claras sobre o que é ciência de dados e o que é engenharia de dados, sendo uma influência decisiva para o início de nosso trabalho de pesquisa.

Lee (LEE, 2020) explora estratégias de big data aplicadas a governos, sociedade e formulação de políticas públicas, destacando oportunidades e desafios na adoção de tecnologias de análise de dados em larga escala. Por meio de estudos de caso, o autor apresenta iniciativas como cidades inteligentes e aplicações em saúde, segurança pública e desenvolvimento sustentável, enfatizando como a análise de big data permite maior eficiência, transparência e responsividade na administração pública. Apesar do potencial, desafios relacionados à proteção de dados e desigualdades de acesso são destacados, exigindo uma abordagem ética e inclusiva. As recomendações incluem parcerias público-privadas e infraestrutura tecnológica adequada para fomentar uma governança mais orientada por dados e centrada no cidadão.

Para expandir os conceitos-chave de gerenciamento de dados (GD), Grillenberg et al. (GRILLENBERGER; ROMEIKE, 2017) usaram uma abordagem empírica, baseada no trabalho de Denning(DENNING, 2003), que consistiu em gerar um sistema de categorias, analisando os 6 livros mais usados no ensino de GD, e outros livros com assuntos mais atualizados sobre GD. A categorização dos conceitos foi feita durante o processo de análise da literatura, com todos os termos relevantes sendo inicialmente listados, seguidos da remoção de termos muito abrangentes ou muito detalhados. Durante o desenvolvimento do sistema de categorias, os termos foram selecionados a partir do livro mais abrangente, seguido dos outros, e então, manualmente agrupados. Para validar os termos encontrados no levantamento da literatura, foram analisados, de forma programática, 305 documentos, totalizando 9447 páginas e slides, resultando em 229 termos mencionados, pelo menos, 300 vezes. O modelo final foi baseado no arcabouço proposto por Denning (DENNING, 2003): tecnologias centrais, práticas, princípios de design e mecânicas.

A partir desse processo, um modelo com 34 conceitos-chave foi gerado, facilitando a inclusão de novos tópicos para ensino e a validação dos tópicos ensinados atualmente. No nosso trabalho utilizamos uma abordagem semelhante para criar as taxonomias de tópicos das bibliografias e ementas, porém, analisamos uma base bem maior de livros, com 373 entradas e 108 ementas. Usamos o modelo gerado pelo trabalho de (GRILLENBERGER; ROMEIKE, 2017) para validar as taxonomias geradas, inclusive a da indústria.

A ACM também notou tais mudanças na área de gerenciamento de dados, tendo em seu relatório de 2020 (ACM and IEEE, 2020), enfatizado a importância do envolvimento da indústria na formulação de competências para o trabalho e a necessidade de conselheiros consultivos do mercado para o desenvolvimento de um currículo significativo. A Ciência de Dados (CD), um novo campo da computação intimamente relacionado com a análise de dados e a engenharia de dados, é um exemplo da necessidade de uma abordagem de ensino atualizada. E em seu novo currículo de computação (ACM; AAAI, 2024) a ACM destaca que, hoje, espera-se que os graduados tenham o conhecimento de um usuário, e não de um desenvolvedor. Isso implica uma maior ênfase no aprendizado de construção de consultas e modelagem de dados. Além disso, é vital que os alunos estejam cientes das tecnologias emergentes, como NoSQL, bancos de dados em nuvem, MapReduce e Dataframes. O documento também ressalta que existe uma tensão entre o foco curricular na preparação profissional e o estudo de uma área de conhecimento como cientista, que é especialmente palpável em gerenciamento de dados. Em nossa pesquisa, usamos o currículo de computação da ACM para complementar nossa taxonomia de tópicos para a indústria.

Para suprir a necessidade de profissionais para lidar com tarefas de análises de dados, algumas iniciativas estão acontecendo no sentido de se criar uma nova área chamada “Ciência de Dados”. A ACM (ACM, 2021) e a SBC (SBC, 2021), definiram currículos para criação de um curso de ciência de dados, que visa formar profissionais capazes de suprir a demanda da indústria quanto à análise do volume exorbitante de dados gerados a cada ano. Em (ACM, 2021), a ACM define que Ciência de Dados é uma área interdisciplinar que combina o domínio dos dados, a ciência da computação e a estatística para interrogar os dados e extrair informações úteis. Cientistas da computação desempenham um papel crucial nesse campo, trazendo métodos para armazenar, proteger a privacidade e integrar dados. Além disso, eles fornecem expertise em aplicar computação de alto desempenho em sistemas distribuídos de maneira eficiente e oferecem ferramentas para analisar e consumir todos os tipos de dados. No entanto,

no próprio relatório, a ACM admite que implementar a Ciência de Dados em um ambiente acadêmico pode apresentar desafios administrativos devido à sua natureza interdisciplinar. Também identificaram, junto à indústria, que características de computação são mais relevantes para pessoas que lidam com dados do que com estatística e matemática.

Por isso, alguns trabalhos verificaram se seria possível formar profissionais com expertise em engenharia de dados nos cursos de Ciência da Computação, sem alterar sua grade significativamente. Em (HASSAN; LIU, 2019), os autores identificaram seis competências essenciais para a CD, incluindo pensamento computacional, pensamento estatístico, fundamentos de matemática, construção e avaliação de modelos, fundamentos de algoritmos e software, e curadoria de dados. Concluíram que tais competências podem ser cobertas perfeitamente pelo currículo atual de Ciência da Computação, com pequenos ajustes em disciplinas específicas para incluir os tópicos relacionados.

Em (SILVA et al., 2014), destaca-se a recente importância do que chamam de "Big Data Management Systems" (BDMS) no panorama atual de bancos de dados, ressalta-se a necessidade de integrar o estudo desses sistemas no currículo de computação, especialmente no âmbito do Gerenciamento de Dados (GD). Propõem a inclusão de novas unidades de aprendizado, como MapReduce (com foco em Hadoop, Dryad e MapReduce), NoSQL (abrangendo Cassandra, BigTable, MongoDB), e NewSQL (por exemplo, VoltDB, Google Spanner), com objetivo de que, ao final do estudo, os alunos reconheçam as principais propriedades, forças e limitações de cada tipo de BDMS; construir aplicações usando esses sistemas e; entender quando é mais adequado utilizar cada um deles.

Essas unidades de aprendizagem propostas podem ser integradas em cursos introdutórios ou avançados. Por exemplo, uma introdução ao MapReduce pode ser incorporada em cursos introdutórios de bancos de dados ou sistemas distribuídos. Da mesma forma, uma visão geral de alto nível sobre NoSQL e NewSQL pode ser incluída em um curso introdutório de banco de dados, usando uma tarefa focada na exploração de tecnologias de gerenciamento de dados além dos bancos de dados relacionais. Os autores do artigo avaliaram a integração de três módulos propostos e a eficácia dos objetivos de aprendizagem por meio de uma pesquisa com uma turma avançada de uma disciplina de bancos de dados. Apesar do tamanho reduzido da turma, os resultados foram, de acordo com eles, promissores, com distribuições de pontuação semelhantes para os módulos e a maioria dos alunos reconhecendo as propriedades de tecnologias como MapReduce, NoSQL e NewSQL e a utilidade de sistemas como Hadoop, HBase e VoltDB.

Além disso, os alunos valorizaram a utilização de exercícios práticos e diagramas ilustrativos. Diante da crescente demanda por processamento escalável e distribuído de grandes conjuntos de dados, os chamados BDMS são amplamente utilizados, e os formandos frequentemente encontram oportunidades de emprego em empresas que utilizam ou planejam utilizar esses sistemas para extrair conhecimentos a partir de seus dados. Portanto, é crucial integrar o estudo desses sistemas nos currículos de computação.

Em um estudo recente, Burke et al. (BURKE; BAILEY, 2019) apresentam os resultados de uma pesquisa com 50 alunos de faculdades e bootcamps, nos EUA, para entender esse novo modelo de cursos rápidos e intensos. Eles levantaram dados demográficos e verificaram como foi o desempenho dos alunos em relação ao mercado de trabalho, em comparação com a educação formal. Foi identificado que os alunos de bootcamps são pessoas mais velhas, entre 20 e 50 anos, a maioria com alguma graduação (86,5%), e com média de 6 anos de experiência no mercado de trabalho. Eles são focados em encontrar um emprego ou conseguir uma promoção. Foi ainda reportado que o ambiente do bootcamp é bastante dinâmico e desafiador, pois o conteúdo é intenso e muito prático, tomando muito tempo dos alunos e exigindo uma dedicação grande.

Na universidade, 21% dos alunos entrevistados reportaram estar trabalhando em projetos junto à indústria, enquanto, no bootcamp, esse número foi de 77%. Esses resultados sugerem que os bootcamps proporcionam uma ênfase mais imediata em termos de resultados práticos, colaboração com a indústria, aquisição de emprego e experiência prática do que os programas de graduação. Isso não significa que as graduações de ciência da computação universitárias careçam desse conteúdo, mas sim, que os estudantes relataram que isso ocorre mais tarde, geralmente durante o último ano da graduação. Já os estudantes de bootcamp relataram apreciar principalmente a natureza acelerada e imersiva de seus programas, no entanto, a partir de suas respostas, não fica claro em que medida essa estrutura e esse currículo incutiram suas habilidades de resolução de problemas e colaboração em comparação com simplesmente aprimorá-las como habilidades já existentes. Esse artigo contribuiu para a compreensão das diferenças entre os dois modelos educacionais e seus impactos na preparação para o mercado de trabalho.

Em outro estudo, Yu Cheng et al.(TU et al., 2018) descrevem os resultados de um programa de pós-graduação no estilo de bootcamp para inclusão de pessoas na área de Tecnologia, financiado pelo governo da Nova Zelândia. Com conteúdos acadêmicos, como seminários, laboratório, aulas, e conteúdos da indústria, a exemplo, os seminários dados por empresas e hackatons, e, nas últimas duas semanas, um projeto final com um problema real da indústria, o

programa conseguiu resultados interessantes. A maioria dos alunos continuou sua formação e/ou conseguiu um emprego na área. Foi identificado que esse modelo traz uma carga de trabalho para alunos e professores maior que um programa tradicional, além disso, é difícil manter o curso atualizado e encontrar laboratórios para todos os alunos. Outras iniciativas semelhantes, em universidades norte-americanas, com resultados similares, foram apresentadas por Hickey et al. (HICKEY; SALAS, 2013).

Por sua vez, Shebaro et al. (SHEBARO, 2018), exploraram uma estratégia de ensino para cursos introdutórios de banco de dados, incorporando pedagogias de aprendizado ativo. Eles enfatizam a necessidade de envolver universitários em autoaprendizagem, atividades em equipe e habilidades de pensamento de ordem superior, afastando-se dos métodos tradicionais baseados em aulas estáticas. A abordagem inclui estudos de caso em sala de aula e diversos papéis estudantis para aumentar o engajamento e a compreensão prática. Os autores demonstram o impacto positivo dessa metodologia no desempenho dos alunos, defendendo sua adequação na educação introdutória de banco de dados.

Em outro estudo, Yang Wang et al. (WANG; MCCOEY; BLUM, 2020) apresentam um modelo inovador de ensino de bancos de dados para calouros em Ciência da Computação e Tecnologia da Informação em universidades norte-americanas. O modelo propõe dois caminhos paralelos de aprendizado: um teórico e outro prático. Esse método foca na introdução precoce de SQL e conceitos abstratos de banco de dados, combinando teoria e prática de maneira integrada e sequencial. Com esse modelo, os autores buscam superar desafios como a limitada experiência prévia dos alunos e a complexidade do conteúdo, proporcionando um aprendizado mais efetivo e engajador. Os resultados com o método foi satisfatório, alcançando 92,5% da nota máxima (4.692/5), sendo mais da metade dos alunos tendo achado o curso “muito valioso”. Esse artigo é fundamental para entender abordagens contemporâneas de ensino de bancos de dados.

Avaliando intervenções em um curso de bancos de dados relacionais, Boisvert (BOISVERT, 2019) analisa as mudanças implementadas ao longo de seis anos em um módulo de bancos de dados relacional. O autor avalia diversas intervenções pedagógicas e tecnológicas, constatando que a maioria não resultou em melhorias significativas no aprendizado dos alunos, contrariando as expectativas. Ele destaca que fatores não abordados, como o engajamento dos alunos, tiveram maior influência nos resultados do que as mudanças implementadas. Esse estudo oferece uma percepção crítica sobre as dificuldades de avaliar intervenções educacionais em ambientes práticos e sobre o impacto de introduzir novas tecnologias de mídia para o ensino.

É importante salientar que o autor deste trabalho de mestrado atua na indústria há 14 anos e como fundador e instrutor em um *bootcamp*, voltado para a área de dados há 6 anos; tendo treinado mais de 1000 alunos neste período, tanto em cursos avulsos, como em cursos para empresas. A motivação da fundação deste *bootcamp* foi a dificuldade de se contratar pessoas para engenharia de dados e análise de dados (AD), e a baixa oferta de cursos neste modelo no Brasil à época. Em tal curso, com duração de 48h, divididos em 4 dias, em que o foco era apresentar, de forma prática, as principais ferramentas e técnicas usadas tanto para ED, quanto para AD.

Os alunos assistiam à aula, copiando os códigos que os instrutores apresentavam, e eram incentivados a experimentar, por alguns minutos, e a fazer perguntas em relação ao conteúdo apresentado. Os instrutores tinham uma aula pré-definida, mas tinham a liberdade para mostrar técnicas relacionadas durante a aula. Depois de cada dia de aula, os alunos tinham exercícios para executar, em um ambiente interativo (*jupyter notebook*). Ao final do curso, os alunos relataram que a intensidade foi alta, porém, conseguiram absorver o conteúdo, e relataram também que o curso serviu para entender mais das áreas de ED e AD, e também auxiliou na migração para essas áreas de atuação.

Em Guolla (GUOLLA, 1999) o autor explora a relação entre a qualidade do ensino e a satisfação dos estudantes, utilizando como base um framework de satisfação do cliente aplicado ao contexto educacional. O estudo, conduzido com estudantes de graduação e MBA, identifica fatores-chave de qualidade no ensino — como aprendizado, entusiasmo do instrutor e organização do curso — e seus impactos variados na satisfação com o curso e com o instrutor. O método de Guolla integra instrumentos consolidados da psicologia educacional com princípios da pesquisa em marketing, oferecendo uma abordagem diagnóstica para aprimorar a eficácia do ensino. Este trabalho conecta as áreas de marketing e educação ao adotar perspectivas multidimensionais sobre qualidade de serviço e satisfação, destacando descobertas acionáveis para instrutores melhorarem os resultados educacionais. Nos inspiramos no modelo de questionário para avaliar a satisfação dos alunos quanto à disciplina, no que tange o conteúdo, sua aplicabilidade no dia a dia, e o método de ensino.

Em Douglas (DOUGLAS, 2006) os autores analisam três métodos para avaliação da qualidade do ensino em instituições de ensino superior: questionários de feedback dos estudantes, observação por pares e o uso de “estudantes misteriosos”. O estudo destaca vantagens e limitações de cada abordagem, como a subjetividade presente nos feedbacks estudantis, o efeito

da observação direta sobre o comportamento dos professores e a viabilidade prática do uso de estudantes misteriosos. A pesquisa defende a triangulação desses métodos para garantir maior validade e confiabilidade na avaliação, promovendo uma cultura de crítica construtiva que priorize o aprimoramento da qualidade do ensino, em alinhamento com os padrões exigidos por agências de acreditação e financiadores. Utilizamos esse artigo para entender possíveis lacunas na avaliação utilizando questionários e como mitigá-las.

O formulário para pesquisa junto aos alunos e à indústria foi criado a partir de dois formulários. O primeiro usado em uma pesquisa sobre uma nova abordagem de ensino de computação para alunos do Ensino Médio (MONTEIRO; FALCãO; RODRIGUES, 2024); e o segundo, um formulário de avaliação de disciplinas da UFJF[1].

Diante desses trabalhos, decidimos concentrar nossa pesquisa em engenharia de dados, conforme definido no Capítulo 2. Além disso, identificamos que as habilidades requeridas para um engenheiro de dados são totalmente abrangidas pelos currículos típicos dos cursos de Ciência da Computação. Pois isso nos permitiria formar profissionais de forma eficaz e rápida, sem a necessidade de criar um novo curso de graduação, evitando a burocracia e custos associados.

[1] Formulário de Avaliação de Disciplina

4

Criação da taxonomia da academia e da indústria

4.1 Levantamento das Bibliografias e Ementas das Disciplinas

Nesta seção, apresentamos pesquisa feita sobre a bibliografia e as ementas utilizadas nas disciplinas de Engenharia de Dados (ED). A intenção é compreender as tendências, práticas e, potencialmente, lacunas existentes na estruturação dos currículos de ED. A pesquisa incluiu um levantamento de livros de referência, a criação de uma taxonomia de tópicos abordados e uma análise das ementas das disciplinas. A investigação também procurou identificar a frequência com que diferentes tópicos são abordados nas ementas, visando revelar quais tópicos são mais enfatizados e quais podem ser negligenciados.

O processo de levantamento da bibliografia, descrito no Capítulo 2, resultou em uma lista inicial de 373 livros. Após filtragem por disciplinas relacionadas à ED, a lista foi reduzida para 250 e, em seguida, foram removidas duplicatas, gerando uma lista final de 80 livros. Dessas referências, 10 não tinham informações de sumário disponíveis. A partir dessa bibliografia, foi criada uma taxonomia de tópicos, resultando em 245 tópicos com três níveis de profundidade. Os tópicos de cada nível variaram de genéricos a específicos. Essa taxonomia completa pode ser encontrada no Apêndice F.

Na Figura 1, apresentamos uma parte dessa taxonomia com os tópicos somente até o segundo nível.

Embora a bibliografia listada nas disciplinas de ED não represente todo o conteúdo

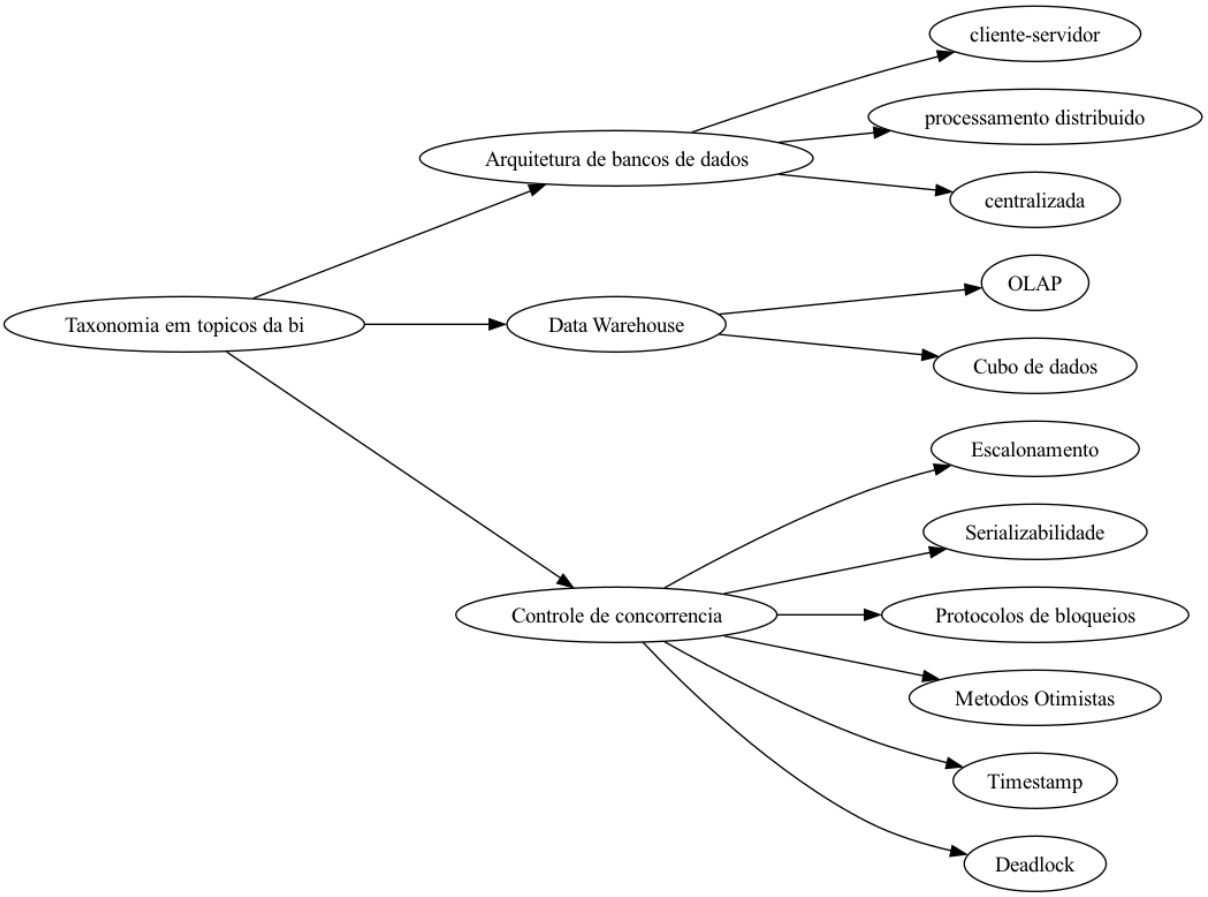


Figura 1 – Amostra da taxonomia de tópicos da bibliografia.

ensinado em sala de aula, consideramos que a taxonomia resultante fornece uma base para todo o trabalho e representa o universo de assuntos que poderiam ser abordados nas disciplinas. Foi possível observar que a maioria das universidades usa os mesmos livros de referência, reduzindo o universo de 250 para 80 livros (68% de redução). Alguns livros e autores dominam a bibliografia, com os top 5 livros em número de aparições nas bibliografias, representando 46% de toda a lista.

A partir das ementas das disciplinas selecionadas, foi criada uma taxonomia de tópicos com 138 tópicos, sendo esse um subconjunto da taxonomia de tópicos da bibliografia. Foi feita uma medida de frequência da ocorrência do tópico nas ementas para entender a quantidade de vezes que cada tópico era mencionado. A taxonomia completa com as frequências dos tópicos pode ser vista no Apêndice G. Na Figura 2, apresentamos uma parte dessa taxonomia com os tópicos somente até o segundo nível.

Com base na frequência de citações dos tópicos nas ementas, podemos observar que, como esperado, os tópicos relacionados a bancos de dados relacionais são amplamente abor-

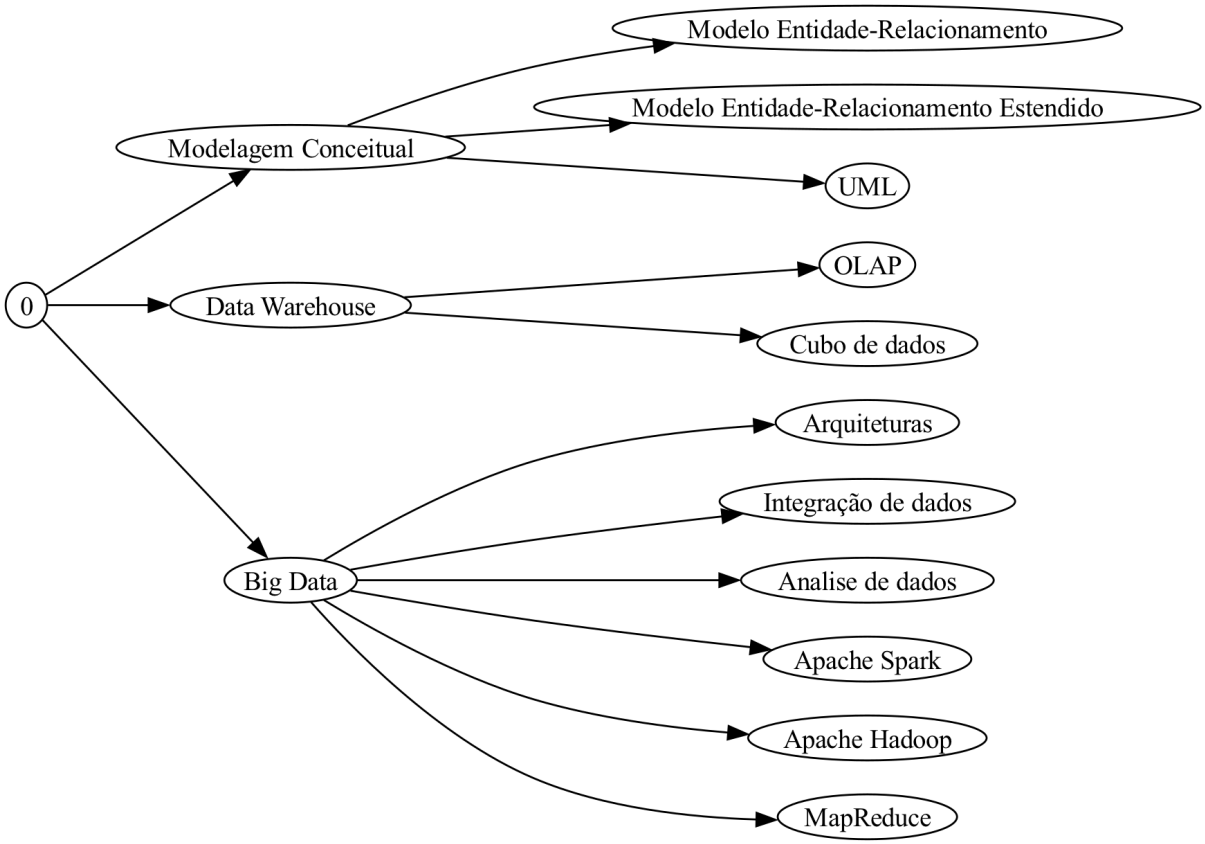


Figura 2 – Amostra da taxonomia de tópicos das ementas.

dados. Tópicos como “SQL”, “Modelo Relacional”, “Formas Normais”, “Álgebra Relacional”, “Restrições”, “Gatilhos”, “Projeto Físico”, “Processamento e Otimização de Consultas”, “Controle de Concorrência”, “Transações”, “Gerenciamento de Bancos de Dados Relacionais”, “Arquitetura de Bancos de Dados” e “Modelo Objeto-Relacional” estão entre os mais mencionados, destacando a importância desses temas no ensino de bancos de dados atualmente. Por outro lado, alguns tópicos considerados em desuso atualmente, como modelos de dados hierárquicos e em rede, bancos de dados heterogêneos, XML e bancos de dados orientados a objeto, ainda ocupam espaço nas ementas. Não podemos afirmar como tais tópicos são ensinados, se são usados para contextualização histórica ou ainda como tecnologias a serem utilizadas no dia a dia.

Uma observação possível de fazer, logo depois desse levantamento, foi a baixa frequência de citações para tópicos mais recentes, como “Processamento de Dados Massivos”, “Big Data”, “Data Streaming”, “Cloud Computing” e “Bancos de Dados Paralelos”, todos com uma citação, entre todas as ementas selecionadas.

4.2 Análise das necessidades da indústria brasileira em relação à Engenharia de Dados

Nesta seção, discutiremos os resultados de uma pesquisa realizada junto a líderes de empresas de tecnologia brasileiras, cujo objetivo foi identificar as necessidades específicas dessas empresas em relação aos engenheiros de dados. Além disso, faremos uma análise detalhada das respostas obtidas. Com base nessa análise, desenvolvemos uma taxonomia de tópicos, que pode ser vista no Apêndice I.

Para entender a realidade atual da indústria brasileira e suas necessidades quanto à engenharia de dados, fizemos levantamento com dez pessoas-chave, em cargos de liderança, na área de tecnologia de várias empresas relevantes no Brasil. Como base na experiência dos próprios autores na indústria, partimos de uma lista de 14 habilidades que consideramos relevantes para engenheiros de dados dos recém-graduados.

Em cada uma dessas habilidades, demos 3 opções para definir o nível de conhecimento desejável em cada uma: Conhecimento Teórico, Conhecimento Prático Básico e Conhecimento Prático Avançado (básico, intermediário e avançado). Também deixamos uma pergunta discursiva para detalhar as escolhas feitas na primeira pergunta e uma terceira pergunta para listar habilidades que não estavam incluídas na lista da primeira pergunta, mas que os participantes do levantamento achavam relevantes de serem mencionadas. Note-se que o objetivo da pesquisa não foi cobrir toda a indústria brasileira, mas sim, termos uma noção inicial do que é esperado pela indústria. Com as respostas desse levantamento, foi possível iniciar a construção de uma taxonomia de tópicos, que, como esperado, é um subconjunto da taxonomia de tópicos das ementas e tópicos recentes necessários para a realidade atual, citados nas respostas. O resultado do levantamento pode ser visto na Tabela 1.

Os resultados do levantamento indicaram que o conhecimento em SQL é uma habilidade importante, com 60% dos entrevistados relatando ser necessário conhecimento prático básico e 20% relatando ser necessário conhecimento prático avançado. A modelagem de dados relacionais também foi considerada relevante, com 70% dos entrevistados relatando ser necessário conhecimento prático básico. O conhecimento em Bancos de Dados NoSQL foi considerada uma habilidade para o qual é necessário conhecimento teórico por 70% dos entrevistados. Já o conhecimento sobre Criação e Gerenciamento de ETLs[1] tem 80% das respostas indicando a ne-

[1] abordagem de processamento de dados no qual os dados são extraídos de várias fontes, carregados em um ambiente de armazenamento e, em seguida, transformados e preparados para análise e uso posterior.

Tabela 1 – Levantamento das habilidades esperadas dos engenheiros de dados.

Habilidade	Teórico	Prático Básico	Prático Avançado
SQL	20%	60%	20%
Modelagem de Dados Relacionais	30%	70%	0%
NoSQL Databases	70%	20%	10%
Criação e gerenciamento de ETLs	0%	80%	20%
Streaming Processing	50%	40%	10%
Batch Processing	10%	80%	10%
Distributed Computing	60%	30%	10%
Distributed File System	70%	20%	10%
Cloud Computing	20%	80%	0%
Arquitetura de Data Lake	60%	20%	20%
Arquitetura de Data Warehouse	50%	30%	20%
Arquitetura de Data Lakehouse	70%	10%	20%
Programação	0%	40%	60%
Automatização de Infraestrutura	50%	50%	0%

cessidade de conhecimento prático básico, mostrando uma clara necessidade de que engenheiros de dados saibam usar ferramentas e desenvolver pipelines de dados.

Batch processing e cloud computing tem 80% das respostas indicando a necessidade de conhecimento prático básico, mostrando que tratar dados em lote é a principal técnica esperada de um engenheiro de dados, em comparação ao processamento de dados via Streaming, e que o trabalho no dia a dia será feito em um ambiente de Cloud (pública ou privada) e que esse domínio é relevante. Esse fato é corroborado quando vemos que para Automatização de Infraestrutura foi indicado como necessidade de conhecimento básico em 50% dos casos. Isso pode ser explicado por estar se tornando cada dia mais comum que os engenheiros cuidem de suas ferramentas nesses ambientes, logo, eles devem ser responsáveis por configurar, disponibilizar e otimizar seu uso. A expectativa sobre o conhecimento de arquiteturas de Data Lake e Data Warehouse ficou muito equilibrada, sendo esperado muito conhecimento teórico e menos prático.

O último ponto interessante que queremos destacar é a relevância de habilidade em programação no levantamento, tendo 40% de respostas em conhecimento prático básico, e 60% em conhecimento prático avançado. Concluímos que tal relevância se dá, pois a construção de pipelines ETLs é feita por meio de programação, utilizando principalmente Python e Scala. Mesmo ferramentas "no code" ou "low code"[2] precisam de customizações e as tarefas de processamento de dados se dão por meio de códigos mais elaborados que o SQL é capaz, principalmente com Apache Spark, e, por fim, também é exigido desse engenheiro que ele construa algoritmos, APIs e programas auxiliares para tarefas do dia a dia.

[2] Plataformas de desenvolvimento que permitem criar aplicativos ou sistemas sem a necessidade de escrever código complexo, utilizando interfaces gráficas e recursos visuais intuitivos.

A taxonomia da indústria, que pode ser vista no Apêndice I, tem como principais diferenças da taxonomia de ementas o foco maior em NoSQL na parte teórica, com tópicos específicos para bancos de dados dessa categoria, como Tipos de Bancos, Projeto Físico, Gerenciamento de Bancos de Dados NoSQL, entre outros. Isso ocorre porque tais bancos de dados têm diferenças cruciais dos bancos de dados relacionais. Tais tópicos foram tirados do (ACM; AAAI, 2024).

Também foram reorganizados os subtópicos de bancos de dados relacionais, agrupando Projeto Físico, Bancos de dados distribuídos, Otimização de consulta e Gerenciamento de Bancos de Dados, no tópico abaixo, pois queremos destacar que tais subtópicos dizem respeito somente ao universo de dados relacionais, e por isso, não podem ocupar o primeiro nível da taxonomia. Ao tópico de Mineração de Dados, foram adicionados os subtópicos de Processamento em Lote, Streaming, ETLs, Ingestão de Dados e Transformação, pois, segundo (BIE et al., 2022), esse trabalho de minerar dados tem tais tarefas e tecnologias envolvidas. Em Big Data, foram adicionados os tópicos de Data Lake e Data Lakehouse como especificidade de Arquitetura, dado que a primeira é estabelecida na indústria como padrão; e a segunda é um tópico emergente (ZAHARIA et al., 2021).

Ainda foram adicionados ao tópico de Big Data os subtópicos de Bancos de dados NewSQL (SILVA et al., 2014), Governança de Dados, Armazenamento e Computação distribuída e Evolução de Esquema, por serem assuntos vitais para o armazenamento e processamento de grandes massas de dados. Em Cloud Computing, foram adicionados os tópicos de Automatização e Infraestrutura, pois se revelou importante no levantamento feito junto à indústria, Contêineres, visto que, hoje, é o principal meio para rodar aplicações em ambientes de cloud computing, Segurança, Custos, Monitoramento e Técnicas de SRE. Em Modelagem Conceitual, foram adicionados subtópicos de modelagem de dados para bancos de dados não relacionais, como bancos de dados de documentos e grafos. Na Figura 3, apresentamos uma parte dessa taxonomia com os tópicos somente até o segundo nível.

4.3 Discussão

Nesta seção, com base nos levantamentos apresentados nas seções anteriores, apresentamos um panorama da situação atual do ensino de Engenharia de Dados nas universidades brasileiras. E ainda um panorama da cobertura dos tópicos ensinados frente às necessidades da indústria brasileira por profissionais formados nessas universidades.

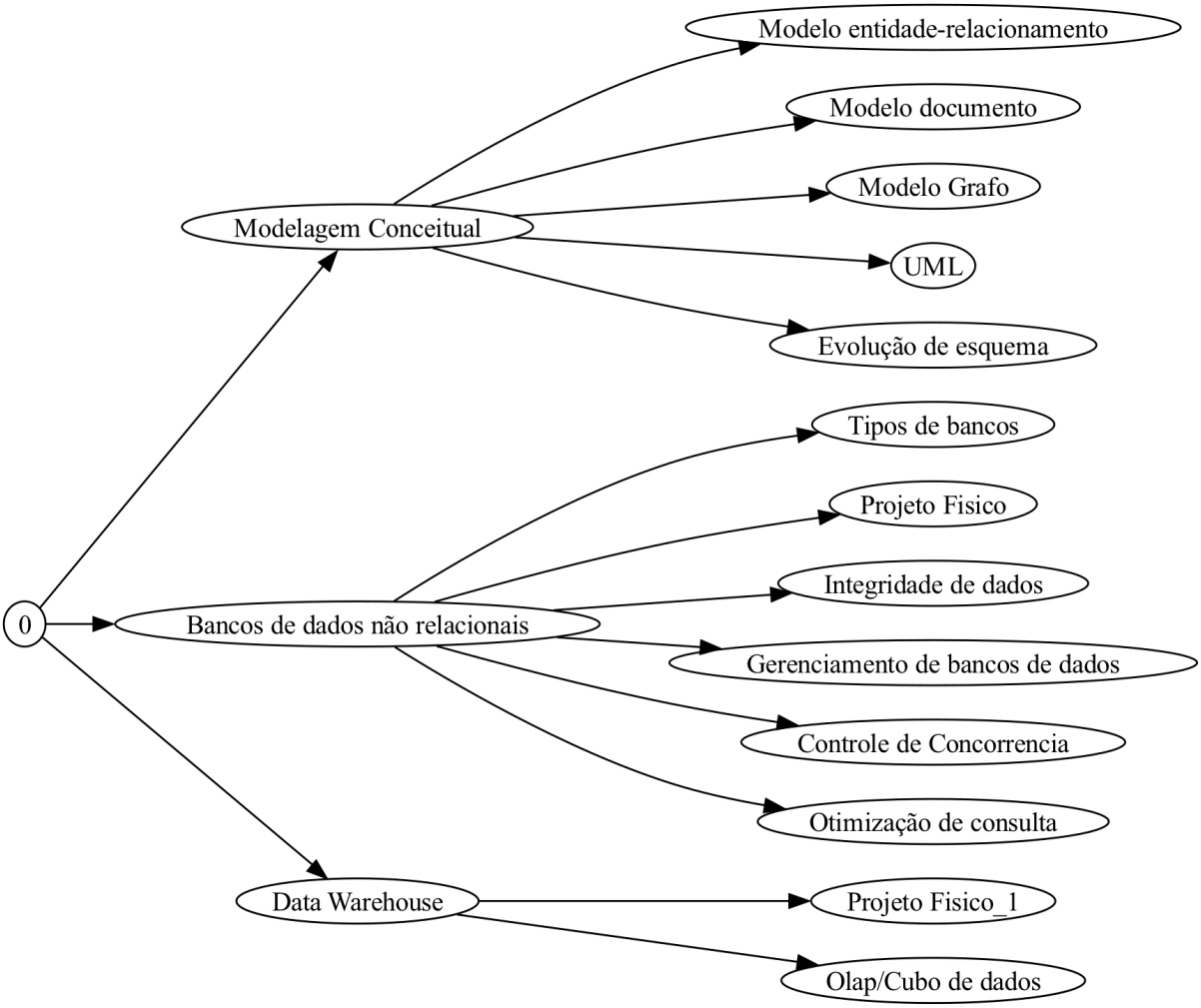


Figura 3 – Amostra da taxonomia de tópicos da indústria.

Inicialmente, pode-se notar que a taxonomia das ementas (Seção 4.1) tem 138 tópicos, enquanto a da indústria (Seção 4.2) é um pouco menor, com 79 tópicos. Isso ocorre principalmente porque as ementas trazem muitos tópicos explicitamente, enquanto nas da indústria, esses tópicos podem ter perdido seu destaque, e/ou terem sido agrupados em termos mais genéricos. Um exemplo é o tópico Gerenciamento de Bancos de Dados, que, na primeira taxonomia, está no primeiro nível, com 5 subtópicos e, na segunda taxonomia, é um subtópico de Bancos de Dados Relacionais e Bancos de Dados Não Relacionais. Isso se deu pela pesquisa com a indústria indicar que não seria necessário um aprofundamento nesse tópico. Também acontece com outros tópicos como “Controle de Concorrência”, “Processamento e Otimização de Consultas”, entre outros.

A importância do conhecimento teórico e prático em bancos de dados não relacionais fica evidente na pesquisa feita junto à indústria, e é um tópico fundamental atualmente, pois tanto

na academia, quanto na indústria, cada vez mais se considera que somente bancos de dados relacionais não são mais suficientes para lidar com a quantidade e a variedade de dados com as quais é necessário lidar hoje em dia (GRILLENBERGER; ROMEIKE, 2014; STONEBRAKER; ÇETINTEMEL, 2005; SILVA et al., 2014). Por isso, expandimos o tópico listado no primeiro nível da taxonomia das ementas para detalhar e deixar evidente que, conforme o levantamento junto à indústria, esses tipos bancos de dados deve ser estudado numa profundidade semelhante aos tradicionais bancos de dados relacionais.

O tópico "Modelagem Conceitual"está presente nas duas taxonomias e sua importância não desapareceu, no levantamento junto à indústria. Esse tópico teve destaque para 70% dos respondentes como importante habilidade prática. Tal resposta se dá pela prevalência dos bancos de dados relacionais no mercado, e também pela sua importância na construção de um Armazém de Dados.

O tópico "Mineração de Dados"passou a incluir os métodos de Processamento de dados em lote e via Streaming, que serão usados para Ingestão de Dados, Transformação de Dados, e garantir a Qualidade dos dados. Esses subtópicos representam o trabalho do Engenheiro de Dados no dia a dia e são essenciais para manipulação dos dados (BIE et al., 2022). Na taxonomia das ementas, Mineração de Dados ainda trata de técnicas específicas para tratamento dos dados, como KDD, detecção de outlier e tratamento de dados faltantes (missing values). Esses subtópicos ainda fazem parte do tópico, mas agora estão incluídos nos termos citados acima.

O tópico "Big Data"inclui todos os tópicos da taxonomia de ementas, com a adição de tópicos relevantes para a indústria. O subtópico de Arquitetura recebe dois sub-subtópicos, data lake e data lakehouse, o primeiro por ser a segunda geração de plataformas de dados; e o segundo por ser a evolução dessa plataforma (ZAHARIA et al., 2021). Foi adicionado também um subtópico de Armazenamento Distribuído, pois atualmente esse é o padrão da indústria para Big Data. Um novo tipo de bancos de dados foi incluído, os Bancos NewSQL, os quais são SGBDs para Big Data com características relacionais. Os subtópicos "Governança de Dados"e "Evolução de Esquema"foram incluídos, por serem essenciais em ambientes de Big Data para, mantê-los coesos ao longo do tempo.

O tópico "Data Warehouse"se mantém relevante entre as taxonomias, dada a sua importância para as análises de negócio, usado por praticamente todas as empresas Fortune 500 juntamente com Data Lake. No entanto, essa arquitetura está evoluindo para algo que está sendo

chamado de Data Lakehouse (ZAHARIA et al., 2021). Esse novo conceito foi incluído no tópico Big Data, no terceiro nível, abaixo do tópico Arquitetura. Ainda em Big Data, foi expandido tal tópico para tratar de armazenamento distribuído, tipos de dados estruturado e não estruturado, governança de dados e bancos de dados NewSQL (SILVA et al., 2014).

Durante a pesquisa, ficou evidente que engenharia de dados tem muita aderência à ciência da computação, com tópicos que vão além de lidar com os dados em si, como Computação em Nuvem para processar dados em larga escala, disponibilizar dados para consumo, unir ferramentas por meio de APIs (HASSAN; LIU, 2019), etc. Também foi detectada a necessidade desconhecimento em Computação Distribuída para entender seus conceitos e como podem ser aplicados no contexto de dados.

O tópico de Computação em Nuvem tem uma frequência baixa na taxonomia de ementas, porém, no levantamento junto à indústria ele se mostrou muito relevante, por isso foi mantido e expandido. A Automatização de Infraestrutura (*infrastructure as code*) é um caso a se notar, pois no levantamento com a indústria, ele se mostrou relevante e necessária para criar e manter serviços em plataformas de computação em nuvem e essa tarefa é cada vez mais comum no dia a dia de um engenheiro de dados.

Adicionamos um tópico chamado genericamente de "Programação" para deixar explícito o conhecimento necessário em microsserviços, padrões de projeto e Estrutura de dados. Ficou evidente, na pesquisa, que esse tópico é extremamente relevante para a indústria, quando falamos de engenharia de dados, pois tais engenheiros precisam integrar diversas ferramentas heterogêneas, e, muitas vezes, criar soluções para consumo desses dados via APIs.

Podemos notar que a academia tem um grande foco em bancos de dados relacionais e suas tecnologias, o que pode ser explicado por sua importância, desde a década de 1970. Poucas universidades procuraram agregar tecnologias mais modernas em seu currículo e, quando o fizeram, colocaram-nas com pouco destaque no currículo ou apenas em cursos de pós-graduação, o que acabou gerando uma distância e um descompasso entre a academia e a indústria.

Para mitigar esse descompasso, foi proposta uma disciplina de Engenharia de Dados, conforme exposto no próximo capítulo.

5

Engenharia de Dados – Uma Nova Disciplina

Neste capítulo, apresentamos a proposta de uma nova disciplina a ser ministrada para cursos de graduação em computação como uma estratégia para mitigar as deficiências identificadas no Capítulo 4. Essa proposta utilizou como base os resultados preliminares detalhados em Seção 4.3. Os pré-requisitos obrigatórios para cursar essa disciplina são: conhecimentos em bancos de dados relacionais e domínio em pelo menos uma linguagem de programação. Os alunos serão expostos a conceitos de arquitetura de computadores e computação distribuída, não sendo necessário conhecimento prévio.

A nova disciplina é denominada "Engenharia de Dados" e, no caso específico da UFAM, será oferecida como disciplina eletiva para os alunos avançados dos cursos de Engenharia de Software e Ciência da Computação. É esperado que neste ponto de sua formação os alunos já tenham tido contato com conceitos fundamentais importantes para engenharia de dados e necessários para o dia a dia das atividades de um profissional da área, tais como programação, arquitetura de computadores e computação distribuída, por exemplo. Os alunos também devem ter cursado uma disciplina introdutória em bancos de dados, na UFAM, "Bancos de Dados I", que seria o seu pré-requisito. Deve-se notar, no entanto, que esta mesma disciplina poderia também ser oferecida em outros cursos da área de computação, em diferentes universidades, podendo ser programada para diferentes momentos do curso, conforme as necessidades e o currículo de cada instituição.

5.1 Objetivo Geral

O objetivo da nova disciplina é fornecer aos alunos uma visão teórica aprofundada e habilidades práticas nos seguintes aspectos:

- **Construir base teórica de Engenharia de Dados:** Adquirir conhecimento sólido em engenharia de dados, dominando os conceitos necessários para manipulação e disponibilização de qualquer volume de dados, considerando aspectos de eficácia e eficiência.
- **Aplicar ferramentas de processamento eficiente de dados:** Ganhar proficiência no uso e ambiente de bibliotecas de desenvolvimento de pipelines de dados, aplicadas a projetos reais.
- **Analisar e resolver problemas:** Capacidade de analisar criticamente ferramentas e soluções para problemas que envolvem grandes volumes de dados, incluindo avaliação de desempenho e custos.
- **Construir uma Plataforma de Dados Moderna:** Desenvolver a capacidade de criar plataformas de dados robustas e eficazes, capazes de tratar quaisquer volumes de dados, utilizando ferramentas e técnicas modernas.

A partir desses aspectos, procuramos organizar unidades didáticas de maneira lógica para os alunos poderem adquirir os conhecimentos gradualmente, mesclando teoria e prática e aumentando de complexidade com o avançar das unidades. Ao final do curso, os alunos terão capacidade para entender como computação em nuvem pode ser aplicada na ED, assim como identificar as diferentes arquiteturas de gerações de Plataformas de Dados (Data Warehouse, Data Lake e Data Lakehouse), e ainda entenderão como construir uma plataforma moderna para atender as necessidades atuais da indústria e sociedade. Isso envolve aspectos como a capacidade de definir as camadas da plataforma, escolher o melhor tipo de arquivo para cada objetivo (leitura, escrita), avaliar a utilização de bancos de dados relacionais e não relacionais, utilização de índices, uso de particionamento e *caches*, conforme as necessidades apresentadas. Com uma sólida base teórica e prática, os alunos poderão utilizar as técnicas e arquiteturas mais modernas para criar plataformas de dados, com capacidade de processar quaisquer volumes de dados, atendendo às demandas atuais e futuras. O curso tem como foco preparar o aluno para além das ferramentas atuais, apresentando a teoria e construindo uma base para que ele possa, ao sair da universidade, continuar aprendendo e se atualizando.

5.2 Objetivos específicos

A disciplina tem como objetivos específicos os seguintes:

- Identificar os fundamentos de Engenharia de Dados
- Diferenciar os tipos de banco de dados não relacionais, suas aplicações, vantagens e desvantagens.
- Reconhecer os mecanismos de aquisição de dados de diversas fontes.
- Construir processamento em lote e streaming para diversas fontes de dados.
- Explicar como construir uma plataforma de dados, usando arquiteturas conhecidas de Data Warehouse, Data Lake e Data Lakehouse;
- Definir o funcionamento dos principais tipos de arquivos usados na Engenharia de Dados;
- Projetar construção de fluxos de dados em lote, desde a aquisição até a disponibilização para consumo analítico, passando por transformações de dados aplicadas nas várias camadas da plataforma de dados.

5.3 Definição da Ementa

A ementa proposta para a disciplina é apresentada na Tabela 2. Ela foi definida, baseada nos assuntos listados na Taxonomia da Indústria (Seção 4.2), visando cobrir os tópicos identificados como relevantes neste trabalho. A formulação de tal ementa precisa equilibrar a apresentação da teoria e de princípios fundamentais da área, juntamente com a utilização de ferramentas práticas. Conforme identificado na Seção 4.3, embora seja importante que os alunos adquiram familiaridade com tecnologias amplamente utilizadas, é ainda mais crucial que eles tenham uma compreensão sólida dos fundamentos subjacentes à Engenharia de Dados, permitindo-lhes adaptar-se ao longo de suas carreiras em um campo tecnológico em constante evolução.

Os tópicos de *Dados Massivos (Big Data)* e *Engenharia de Dados* são as bases da disciplina, abrangendo os fundamentos teóricos e as práticas que permitem aos alunos lidar com grandes volumes de dados. Como identificado na pesquisa, estar preparado para lidar com uma quantidade de dados massiva é essencial, e um Engenheiro de Dados deve conseguir fazê-lo. Por isso, esses tópicos perpassam todo o curso, e incluem assuntos como Arquitetura de Dados Massivos,

Armazenamento Distribuído, Tipos de dados, bem como as tecnologias emergentes usadas para processar e armazenar esses dados de forma escalável e distribuída.

No tópico de *Bancos de dados Não Relacionais*, a pesquisa mostrou ser muito importante que os alunos tenham conhecimento teórico neste tipo de bancos dados, visto que eles são necessários para resolver problemas que os bancos relacionais não resolvem tão bem. Por esse motivo, incluímos tópico específico sobre o assunto. Assim, são apresentados diferentes tipos de bancos de dados não relacionais, suas características, aplicações, vantagens e desvantagens. Também são apresentados diferentes algoritmos de indexação e técnicas de particionamento, mostrando seu funcionamento, implementação, e aplicação.

Os tópicos de *ETL* e *Ingestão de Dados* tratam da aquisição de dados, transformação e posterior disponibilização para consumo analítico. A Ingestão de dados trata de apresentar aos alunos os diferentes tipos de fontes de dados que podemos ter, como Bancos de dados Relacionais, Bancos de dados Não Relacionais, APIs, e arquivos estruturados (XML, JSON, CSV, etc.). E dentro deste tópico, temos uma parte dedicada à apresentação da técnica de *Change Data Capture* (CDC), que consiste em adquirir os dados de SGBDs relacionais por meio do seu log de mudança,*Write-Ahead Log* (WAL). *Extract-Transform-Load* (ETL) trata de todo o processo desde a aquisição, a transformação e até a disponibilização do dado para consumo analítico. Trataremos dos conceitos para a construção de um processo eficiente. Eles estão contidos no assunto *Mineração de Dados*, juntamente com outros tópicos importantes que discutiremos mais adiante.

Já no tópico de *Processamento em Lote* dedicamos boa parte a entender sua teoria, suas características, vantagens, desvantagens e aplicações. Dado a sua importância evidenciada na pesquisa, exploramos profundamente o tema, apresentando os algoritmos mais utilizados para distribuição de tarefas em um *cluster* computacional, sincronismo e gerenciamento de erros. Exploramos como a *Computação em Nuvem* se integra nesse ambiente, disponibilizando ferramentas para podermos escalar tais clusters sob demanda.

Para o tópico de *Processamento de Fluxo Contínuo* exploramos suas características, aplicações, algoritmos para habilitar e tratar o fluxo contínuo de dados, arquitetura típica, os desafios, algoritmos de seleção, agregação, projeção e filtragem de dados em tempo real, além de apresentar janelas de processamento e estratégias para lidar com eventos fora de ordem.Serão também explorados casos de uso que demonstrem como esse modelo de processamento pode ser aplicado em sistemas de recomendação e monitoramento de fraudes em tempo real.

O tópico de *Computação em Nuvem* perpassa toda a disciplina, ao estar relacionado a vários outros tópicos. Tratamos de conceitos gerais, de histórico, das necessidades e aprofundamos como funciona um ambiente de Computação em Nuvem, falando da arquitetura de um Data Center, suas necessidades de Rede, Armazenamento e como funciona a Virtualização de Hardware. Este último assunto tem um tópico dedicado, *Virtualização*, por ser muito importante para o processamento de dados massivos de forma distribuída. Nele, está incluído o assunto de contêineres, no qual tratamos dos conceitos e aplicações no dia a dia do Engenheiro de dados.

No tópico de *Arquitetura de Plataforma de Dados*, tratamos das 3 gerações de Plataforma de dados, Armazém de Dados, `Data Lakes` e `Lakehouses`. Temos um contexto histórico da evolução das plataformas, explorando suas características, vantagens e desvantagens. Apresentamos um embasamento teórico profundo para os alunos terem o poder de escolher a melhor solução dado o cenário que estejam atuando. Aqui também é tratado o assunto *Modelagem de Dados* para Armazéns de Dados e Data Lakes.

O assunto *Programação*, incluído na taxonomia (Seção 4.2), não é específico para Engenharia de Dados e assumimos que esse assunto foi extensivamente coberto nas disciplinas básicas dos cursos de graduação em computação. Portanto, não incluímos esse assunto em nenhum tópico de forma explícita. Da mesma forma, o assunto *Bancos de Dados Relacionais*, que foi também incluído na taxonomia, é extensivamente coberto nas disciplinas introdutórias sobre bancos de dados de várias universidades, como visto na Seção 4.3. Portanto, não incluímos esse assunto em nenhum tópico de forma explícita. Também, o assunto de *Estrutura de Dados*, dado em disciplinas anteriores do curso, é usado em partes da disciplina, em que os alunos serão relembrados de conceitos importantes, como árvores, *hashing*, e grafos, essenciais em sistemas distribuídos para garantir eficiência e otimização no armazenamento e consulta de dados.

Tabela 2 – Ementa da nova disciplina de engenharia de dados

Tópicos
Dados Massivos (Big Data)
Engenharia de Dados
Bancos de Dados Não Relacionais
ETL[1], Ingestão de Dados
Processamento em Lote
Processamento em Fluxo Contínuo
Computação em Nuvem, Virtualização
Arquitetura de Plataforma de Dados

5.4 Detalhamento da ementa

Esta seção apresenta um plano de aula da nova disciplina cujo objetivo é fornecer uma referência para eventuais ofertas dessa disciplina, em cursos de computação. Esse plano de aula foi estruturado para fornecer uma base sólida em Engenharia de Dados, abordando desde conceitos fundamentais até as tecnologias mais recentes. O conteúdo foi organizado em uma sequência lógica de unidades, conforme listado na Tabela 3, que visa preparar os alunos para os desafios práticos e teóricos da área.

Para cada unidade da tabela, apresentamos os conceitos teóricos e práticos relevantes, explorando ferramentas amplamente utilizadas no mercado e discutindo suas aplicações em cenários reais. O material de cada unidade foi escolhido para proporcionar uma compreensão aprofundada, permitindo aos alunos dominar as tecnologias e métodos essenciais para a área de Engenharia de Dados. Listamos também o material bibliográfico de suporte, que inclui livros e artigos que auxiliam no aprofundamento dos temas abordados, garantindo uma formação completa e atualizada.

Tabela 3 – Unidades programadas para nova disciplina.

Unidades
Introdução à Engenharia de Dados
Computação em Nuvem
Controle de Versão e Virtualização de Ambientes
Bancos de Dados Não-relacionais
Ingestão de Dados
Processamento de Dados em Clusters
Processamento de Fluxos de Dados Continuo
Plataformas de Dados
ETL e Orquestração

Construímos uma sequência lógica de unidades, listadas na Tabela 3, baseada na ementa definida na seção anterior, que atendam tais necessidades e tragam para os alunos a oportunidade de aprender assuntos relevantes e atuais, ao mesmo tempo que aprofundam seu conhecimento na área de ED.

Introdução à Engenharia de Dados

Iniciamos o semestre com uma contextualização sobre ED, com a unidade *Introdução à Engenharia de Dados*, nela expandimos o universo de gerenciamento de dados com o qual os alunos estão mais familiarizados, mostrando realidades novas, como o volume imenso de dados atualmente disponíveis, o valor do mercado e a multidisciplinaridade da ED. Também são apresentados

os conceitos gerais e fundamentais da ED, e introduzidos os algoritmos mais usados para processamento distribuído, suas características e seus desafios. Para esta unidade, sugerimos dois livros, (ZBURIVSKY D.; PARTNER, 2021) e (KLEPPMANN, 2017). Também foi utilizado o artigo (DEAN; GHEMAWAT, 2004). Ambos foram usados em diversas unidades para apoiar o desenvolvimento das aulas.

- (KLEPPMANN, 2017) KLEPPMANN, M. Designing data-intensive applications: The big ideas behind reliable, scalable, and maintainable systems. Sebastopol, CA, USA: O'Reilly Media, 2017.
- (ZBURIVSKY D.; PARTNER, 2021) ZBURIVSKY, D.; PARTNER, L. Designing cloud data platforms. Nova Iorque, NY, USA: Manning Publications, 2021.
- (DEAN; GHEMAWAT, 2004) Jeffrey Dean and Sanjay Ghemawat. MapReduce: Simplified Data Processing on Large Clusters. OSDI '04: 6th Symposium on Operating Systems Design and Implementation.

Computação em Nuvem

Na sequência introduzimos fundamentos de *Computação em Nuvem*, mostrando como tal ambiente opera e pode contribuir na criação de uma arquitetura de dados robusta. Nessa unidade, tratamos de por que esse assunto é relevante atualmente, como é a estrutura típica de um *Data Center* para servir a um ambiente de nuvem, necessidades de rede e virtualização. Esse último tema é bastante relevante, ao estar diretamente ligado ao tema de Virtualização via contêineres, que perpassa todo o curso e é uma tecnologia muito utilizado no dia a dia de diversas empresas e órgãos da sociedade civil. Recomendamos uma parte específica para falar do *Google Cloud Platform*, que é um provedor público de infraestrutura de nuvem que pode ser utilizada durante todo o curso. Apresentamos os serviços disponíveis e depois analisamos os serviços específicos para Engenharia de Dados. Para esta unidade, sugerimos 4 livros (BARROSO LUIZ ANDRé; HOLZLE, 2013), (ZBURIVSKY D.; PARTNER, 2021), (WIJAYA, 2022), (COSTA R.; HODUN, 2021).

- (BARROSO LUIZ ANDRé; HOLZLE, 2013) Barroso, Luiz André; Holzle, Urs. The Datacenter as a Computer: An Introduction to the Design of Warehouse-Scale Machines, Second Edition. Morgan & Claypool Publishers, 2013.

- (ZBURIVSKY D.; PARTNER, 2021) ZBURIVSKY, D.; PARTNER, L. Designing cloud data platforms. Nova Iorque, NY, USA: Manning Publications, 2021.
- (WIJAYA, 2022) WIJAYA, A. Data Engineering with Google Cloud Platform: A practical guide to operationalizing scalable data analytics systems on GCP. Birmingham, England: Packt Publishing, 2022
- (COSTA R.; HODUN, 2021) COSTA, R.; HODUN, D. Google cloud cookbook: Practical solutions for building and deploying cloud services. Sebastopol, CA, USA: O'Reilly Media, 2021.

Controle de Versão e Virtualização de Ambientes

Incluímos no plano uma breve introdução ao *Controle de Versão e Virtualização de Ambientes*, conceitos e as principais ferramentas utilizadas por profissionais da CC. Para ambas, fazemos uma revisão dos principais comandos e caminhos de utilização, mostrando aos alunos como usar tais ferramentas e porque são importantes. Essa unidade é necessária também para o decorrer da disciplina, pois todos os trabalhos e aulas utilizam ambas as ferramentas como implementar as soluções. Este conhecimento não é relacionado somente à ED, mas a todas as atividades de um profissional de computação, sendo assim, os alunos terão um conhecimento útil para o resto do curso. Nelas, utilizamos Git para controle de versão e Docker para virtualização de ambiente. O foco desta unidade não é explorar profundamente os conceitos de cada tópico, e sim, apresentar aos alunos as ferramentas e prepará-los para usá-las durante o curso. Sendo assim, a bibliografia consiste em manuais de uso prático, que exploram os conceitos básicos, dando aos alunos a possibilidade de aprofundamento nos estudos dessas ferramentas posteriormente (CHACON, 2010) e (VITALINO J. F. N.; CASTRO, 2018).

- (CHACON, 2010) CHACON. Pro Git. Nova Iorque, NY, USA: Springer, 2010.
- (VITALINO J. F. N.; CASTRO, 2018) VITALINO, J. F. N.; CASTRO, M. A. N. Descomplicando o Docker 2a edição. [s.l.] Brasport, 2018.

Bancos de Dados Não Relacionais

Após essas unidades, entramos no universo de *Bancos de Dados Não Relacionais* (NoSQL), em que os alunos são apresentados a novos SGBDs, que não seguem o modelo relacional e têm outros propósitos de uso. Primeiro, fazemos uma introdução ao assunto na totalidade, destacando a necessidade de novos modelos de dados baseados na realidade atual, também fazemos uma revisão do Teorema CAP. Serão apresentados SGBDs NoSQL variados, orientados a documentos, como MongoDB, orientado a colunas para alto desempenho, como Cassandra, chave-valor, como Redis e, por fim, um SGBD base em grafos, como Neo4J. Para esta unidade, foram utilizados os artigos (CATTELL, 2010), (RITTER, 2016), (BREWER, 2012), (DAVOUDIAN, 2018), (CORBELLINI MATEOS, 2017).

- (CATTELL, 2010) Cattell, R., Scalable SQL and NoSQL Data Stores. SIGMOD Record (2010)
- (BREWER, 2012) Brewer, E.A., Pushing the CAP: Strategies for Consistency and Availability. IEEE Computer 45(2): 23-29 (2012)
- (DAVOUDIAN, 2018) Davoudian, Chen & Liu: A Survey on NoSQL Stores. ACM CSUR (2018)
- (RITTER, 2016) Gessert & Ritter. Scalable data management: NoSQL data stores in research and practice. ICDE (2016)
- (CORBELLINI MATEOS, 2017) Corbellini, Mateos, Zunino, Godoy & Schiaffino: Persisting big-data: The NoSQL landscape. Information System (2017)

Ingestão de Dados

Em seguida, enfocamos plataformas de dados modernas, nas quais mostramos como é uma arquitetura de dados para atender às necessidades atuais. Essa arquitetura é apresentada durante várias unidades, visando criar uma solução fim-a-fim para resolver a demanda de processamento de qualquer volume de dados. A primeira unidade apresentada é *Ingestão de Dados*, nela, temos uma introdução ao conceito de Plataforma de Dados na Computação em Nuvem; quais as vantagens e desvantagens e os conceitos teóricos envolvidos. Depois, mostramos os tipos diferentes de ingestão de dados (Incremental, via Change-Data-Capture (CDC), Full Load) e

suas características. Por fim, falamos das diferenças de Ingestão para os diferentes tipos de fontes de dados, como APIs HTTP, Arquivos, Bancos de Dados Relacionais e NoSQL. O livro utilizado para a construção dessa unidade foi o de ZBURIVSKY & PARTNER (ZBURIVSKY D.; PARTNER, 2021).

- (ZBURIVSKY D.; PARTNER, 2021) ZBURIVSKY, D.; PARTNER, L. Designing cloud data platforms. Nova Iorque, NY, USA: Manning Publications, 2021.

Processamento de Dados em Clusters

A próxima unidade trata de apresentar aos alunos os conceitos de *Processamento de Dados em Clusters*. Nessa unidade, os alunos são expostos a novos conceitos e a ferramentas para tal tipo de processamento, explorando as soluções mais modernas existentes, as características necessárias para atingir um processamento distribuído eficiente, desafios encontrados e arquiteturas de solução mais comuns. Destacamos a importância dos serviços de Computação em Nuvem para esse fim, pois com eles é possível o gerenciamento desse ambiente de forma muito mais simples e eficaz. A ferramenta recomendada é Apache Spark, que fornece uma interface programática que abstrai as complexidades de executar processamento de dados paralelamente em um cluster. Para esta unidade, foram utilizados os livros (KLEPPMANN, 2017) e (ZBURIVSKY D.; PARTNER, 2021), (CHAMBERS BILL; ZAHARIA, 2018), e também o artigo (ZAHARIA et al., 2012)

- (KLEPPMANN, 2017) KLEPPMANN, M. Designing data-intensive applications: The big ideas behind reliable, scalable, and maintainable systems. Sebastopol, CA, USA: O'Reilly Media, 2017.
- (ZBURIVSKY D.; PARTNER, 2021) ZBURIVSKY, D.; PARTNER, L. Designing cloud data platforms. Nova Iorque, NY, USA: Manning Publications, 2021.
- (CHAMBERS BILL; ZAHARIA, 2018) Chambers, Bill; Zaharia, Matei. Spark: The Definitive Guide: Big Data Processing Made Simple. O'Reilly Media, 2018.
- (ZAHARIA et al., 2012) Zaharia, Matei and Chowdhury, Mosharaf and Das, Tathagata and Dave, Ankur and Ma, Justin and McCauly, Murphy and Franklin, Michael J and Shenker, Scott and Stoica, Ion. Resilient distributed datasets: A Fault-Tolerant abstraction

for In-Memory cluster computing. 9th USENIX symposium on networked systems design and implementation (NSDI 12), 2012.

Processamento de Fluxo de Dados Contínuos

Temos também uma unidade especial para *Processamento de Fluxo de Dados Contínuos*. Nessa seção, apresentamos os principais pontos de diferença entre o processamento em lote e em fluxo de dados, as características desse tipo de processamento de dados e suas aplicações. São apresentados casos de uso do mundo real, tornando mais tangível a aplicação dos conceitos. Aqui continuamos explorando o Apache Spark, agora com foco em sua interface de *Streaming*, juntamente com o Apache Kafka, uma ferramenta de armazenamento de mensagens e processamento em tempo real, com alta disponibilidade e baixa latência. Para esta unidade, foi utilizado o material previamente desenvolvido para a disciplina em anos anteriores, com complementos do livro (KLEPPMANN, 2017), (ZBURIVSKY D.; PARTNER, 2021), e (CHAMBERS BILL; ZAHARIA, 2018).

- (KLEPPMANN, 2017) KLEPPMANN, M. Designing data-intensive applications: The big ideas behind reliable, scalable, and maintainable systems. Sebastopol, CA, USA: O'Reilly Media, 2017.
- (ZBURIVSKY D.; PARTNER, 2021) ZBURIVSKY, D.; PARTNER, L. Designing cloud data platforms. Nova Iorque, NY, USA: Manning Publications, 2021.
- (CHAMBERS BILL; ZAHARIA, 2018) Chambers, Bill; Zaharia, Matei. Spark: The Definitive Guide: Big Data Processing Made Simple. O'Reilly Media, 2018.

Plataformas de Dados

A penúltima unidade trata de apresentar aos alunos as diferentes arquiteturas de *Plataformas de Dados* em suas 3 gerações: A primeira com Data Warehouses, a segunda com Data Lakes e a terceira, a novíssima arquitetura de Data Lakehouses. Nessa unidade são apresentadas as diferenças entre as gerações, suas principais características e aplicações. Em Data Lakes, é apresentada uma arquitetura em camadas para gravação dos dados, na qual, cada camada tem suas características e necessidades. Também são apresentados os diferentes tipos de arquivos

(CSV, AVRO, Parquet e Delta Lake) usados por Data Lakes e Lakehouse, suas características, vantagens, desvantagens e aplicações. Para desenvolver esta unidade, utilizamos diversos materiais, entre livros, artigos científicos e documentação oficial das ferramentas utilizadas. Nos livros, temos o conceituado KIMBALL&ROSS (KIMBALL R.; ROSS, 2013) e o já mencionado ZBURIVSKY&PARTNER (ZBURIVSKY D.; PARTNER, 2021). Nos artigos, utilizamos aquele que define o conceito de Data Lakehouse (ZAHARIA et al., 2021), e também o artigo que define os conceitos e funcionamento dos arquivos Delta Lake (ARMBRUST, 2020). Por fim, para aprofundar o estudo do formato de arquivo Parquet, utilizamos a documentação oficial, que fornece manuais e informações suficientes para os alunos dominarem os conceitos e aprofundarem os estudos posteriormente.

- (KIMBALL R.; ROSS, 2013) KIMBALL, R.; ROSS, M. The data warehouse toolkit: The definitive guide to dimensional modeling. 3. ed, John Wiley & Sons, 2013.
- (ZBURIVSKY D.; PARTNER, 2021) ZBURIVSKY, D.; PARTNER, L. Designing cloud data platforms. Nova Iorque, NY, USA: Manning Publications, 2021.
- (ZAHARIA et al., 2021) MICHAEL ARMBRUST, ALI GHODSI, REYNOLD XIN, MATEI ZAHARIA. Lakehouse: A New Generation of Open Platforms that Unify Data Warehousing and Advanced Analytics. Proceedings of CIDR. Vol. 8., p. 28, 2021.
- (ARMBRUST, 2020) ARMBRUST, M. et al. Delta lake: High-performance ACID table storage over cloud object stores. Proceedings of the VLDB Endowment International Conference on Very Large Data Bases, v. 13, n. 12, p. 3411–3424, 2020.

ETL e Orquestração

Por último, temos uma unidade para tratar do tema de *ETL e Orquestração*, em que exploramos as necessidades de uma plataforma de dados moderna, como agendamento de tarefas e controle de execução, e também ferramentas para atender essas características. A ferramenta escolhida aqui foi o Airflow, amplamente utilizado pela indústria. Os livros utilizados para a construção dessa unidade foram (ZBURIVSKY D.; PARTNER, 2021), que trazem os conceitos e as teorias por trás de ETLs e Orquestração, e (HARENSLAK B. P.; DE RUITER, 2021), que aprofunda os estudos sobre a ferramenta.

- (ZBURIVSKY D.; PARTNER, 2021) ZBURIVSKY, D.; PARTNER, L. Designing cloud data platforms. Nova Iorque, NY, USA: Manning Publications, 2021.
- (HARENSLAK B. P.; DE RUITER, 2021) HARENSLAK, B. P.; DE RUITER, J. R. Data pipelines with Apache airflow. Nova Iorque, NY, USA: Manning Publications, 2021.

No final da disciplina, os alunos devem poder entender o que é a ED no mundo atual, as necessidades e as habilidades para se trabalhar nesse papel. Também devem conseguir escolher um banco de dados baseado nas necessidades apresentadas, considerando o tipo de dado, de consistência e de replicação. E por fim, devem poder construir uma Plataforma de Dados, ou seja, um conjunto de ferramentas que habilitam o processamento de qualquer volume de dados de forma organizada e eficiente, considerando as necessidades das fontes de dados e dos consumidores desses dados.

Os materiais didáticos desenvolvidos para a disciplina foram slides, que guiaram as aulas, e trabalhos práticos, em que os alunos poderiam praticar o que foi passado em sala de aula. Todos os materiais estão disponíveis nesse repositório[2], sob licença Creative Commons, permitindo sua alteração e reprodução livre.

[2] material didático utilizado

6

Aplicação da Nova Disciplina

Neste capítulo, descrevemos como a disciplina proposta no Capítulo 5 foi ministrada na UFAM. Também apresentamos e discutimos os resultados de uma avaliação feita pelos alunos e pela indústria a respeito da disciplina ministrada. Nesta oferta, a disciplina for ministrada pelo autor e pelo orientador deste trabalho. Ao final, discutimos os resultados encontrados e elencamos algumas sugestões de melhorias para futuras ofertas.

6.1 Planejamento das Aulas e Aplicação no Semestre

O semestre letivo iniciou em 22 de abril de 2024 e terminou em 08 de agosto de 2024, tendo 108 dias corridos. Tivemos diversas interrupções durante o período, como viagens não planejadas e semanas acadêmicas, sendo necessário fazer ajustes nos conteúdos planejados.

Abaixo temos o planejamento das semanas, com a carga horária de cada item, no qual dividimos as unidades para crescerem de complexidade, compondo, no final, uma visão completa de uma plataforma de dados moderna e cobrindo as necessidades atuais identificadas na pesquisa.

- **Semana 1-2:** Introdução à Engenharia de Dados, *Cloud Computing* e Ferramentas (6h)
- **Semana 3-4:** Bancos de Dados NoSQL (8h)
- **Semana 5-6:** Ingestão de Dados (8h)
- **Semana 7-8:** Processamento em Lote e Fluxo de Dados (8h)
- **Semana 9-10:** *Data Lakes* (6h)

- **Semana 11-12:** *Data Warehouses* (10h)
- **Semana 13:** ETL e Orquestração (6h)

Durante as aulas, materiais práticos foram apresentados aos alunos, como demonstrações dos conceitos aplicados a problemas reais, casos de uso de grandes empresas e exibição das ferramentas estudadas.

6.2 Avaliações Teóricas e Trabalhos Práticos

Foram aplicados três Trabalhos Práticos (TP) durante o semestre. No primeiro TP, foi pedido aos alunos para configurarem uma instância do SGBD não relacional MongoDB com alta disponibilidade e fazer nele um carregamento de 500 mil itens dados de reviews de produtos da Amazon, e ainda fizessem um relatório explicando como funciona o mecanismo de alta disponibilidade de ReplicaSet.

O segundo TP compreendeu todas as unidades até Processamento em Lote. Foi pedido para os alunos configurarem dois SGBDS, um MongoDB (reutilizando o primeiro trabalho) e um Postgres, carregando em cada um 5000 itens. Também foi pedido para os alunos criarem um script para consumo de dados de uma API pública. São especificados dois endpoints dessa API. Para as três fontes, foi pedido que os alunos extraiam os dados no formato JSON e depois os alunos devem criar um código de ingestão dos dados que transforma o dado para o formato AVRO. Eles devem estudar o formato AVRO e criar um relatório explicando suas características, vantagens e desvantagens em relação a outros formatos de arquivo.

No terceiro TP, os alunos deveriam implementar uma Plataforma de Dados fim-a-fim, desde a ingestão de dados de bancos relacionais e não relacionais, e uma API pública, até a exibição de um relatório. Eles devem utilizar tudo que foi construído no segundo TP e, a partir daí, desenvolver um código de processamento de leitura dos dados em AVRO, aplicando as seguintes transformações: achatamento das colunas aninhadas, desnormalização dos dados dos SGBDs para gravação em uma única tabela, chamada "Review". Para os dados da API, foi pedido para removerem os dados de 2 colunas, além do achatamento de colunas aninhadas. Esses dados deveriam ser gravados em Parquet. Depois disso, os alunos devem criar um código para reformular os dados da API, criando tabelas novas, baseadas em um critério previamente definido. Para os dados dos SGBDs, os alunos devem criar um código para gerar um relatório

extraindo algumas estatísticas dos dados. Além disso, foi pedido para os alunos que criem um desenho de arquitetura de uma plataforma de dados em um ambiente de Computação em Nuvem público, podendo escolher entre os provedores Google, AWS ou Azure, tanto para processamento em Lote, quanto para processamento via Streaming, incluindo uma análise breve de cada serviço escolhido, suas características e custos para operação. Também neste relatório deve haver uma comparação entre as arquiteturas, listando suas características, vantagens e desvantagens.

Foram realizadas duas avaliações teóricas. A primeira avaliação abrangeu as unidades iniciais, começando com "Introdução à Engenharia de Dados" e seguindo com "Computação em Nuvem", "Controle de Versão e Virtualização de Ambientes", "Bancos de Dados Não Relacionais" até "Ingestão de Dados". A segunda avaliação cobriu as unidades subsequentes, iniciando em "Processamento de Dados em Clusters", passando por "Processamento de Fluxos de Dados Contínuos", "Plataformas de Dados", e finalizando com "ETL e Orquestração".

6.3 Avaliação dos alunos

A turma contou com 40 alunos, e para entender melhor sua demografia, enviamos um questionário, descrito no Capítulo 2, e coletamos informações para entender o perfil dos alunos. Foram coletadas 38 respostas. O formulário completo da pesquisa pode ser visto no Apêndice K.

Foi identificado que a maioria dos alunos está na faixa etária entre 17 e 25 anos, como apresentado na Figura 4. Mais da metade da turma trabalha ou faz estágio em alguma área da computação, porém, menos de 10% trabalham na área de Engenharia de Dados, como pode ser observado na Figura 5.

O formulário de avaliação, descrito no Capítulo 2, foi enviado para os alunos da disciplina no final do período, antes do lançamento das notas. Foi pedido que os alunos respondessem com assertividade e responsabilidade. Também foi garantida a eles uma bonificação na pontuação final para quem respondesse o questionário.

6.3.1 Avaliação do Conteúdo

A primeira parte do formulário foi criada para avaliação do conteúdo da disciplina, quanto ao seu entendimento, multidisciplinaridade e aprendizagem. Os alunos deveriam refletir sobre o

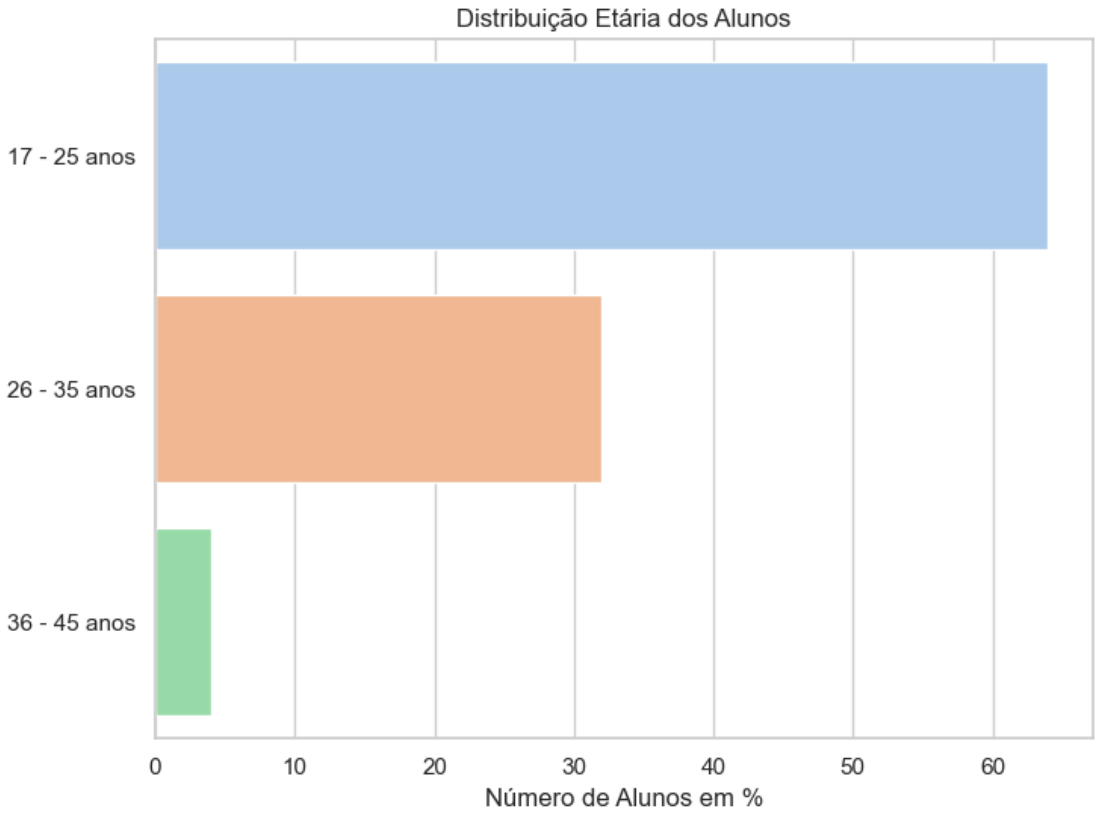


Figura 4 – Distribuição dos Alunos da Turma por Faixa Etária

semestre letivo e responder o quanto eles entendem de cada item perguntado. Os resultados estão na Figura 6.

Na pergunta *Você entende as diferentes gerações de Plataforma de Dados? (DW/Data Lake/Lakehouse)*, 34,21% dos alunos responderam que entendem; 52,63%, que entendem bem e; 13,15%, que entendem completamente. Quando questionados a responder sobre a compreensão das diferenças entre Bancos de Dados Relacionais e Não Relacionais (NoSQL), na pergunta *'Você compreende as diferentes aplicações de bancos de dados Relacionais e NoSQL'*, 55,3% responderam que compreendem completamente; 39,5%, compreendem bem e; 5,3% tem alguma compreensão. Para a afirmação *Na sua opinião, a disciplina abrangeu vários assuntos da computação, como Cloud Computing e Programação*, 71,1% concordam totalmente; 23,7% concordam e; 5,3% estão neutros. E sobre o entendimento de como e por que implementar um processamento em lote ou streaming, 26,3% entendem completamente; 34,2% entendem; 34,2% estão neutros e; 5,3% não entendem.

Foi feita uma pergunta discursiva, para os alunos opinarem livremente sobre a disciplina. Destacam-se alguns comentários:

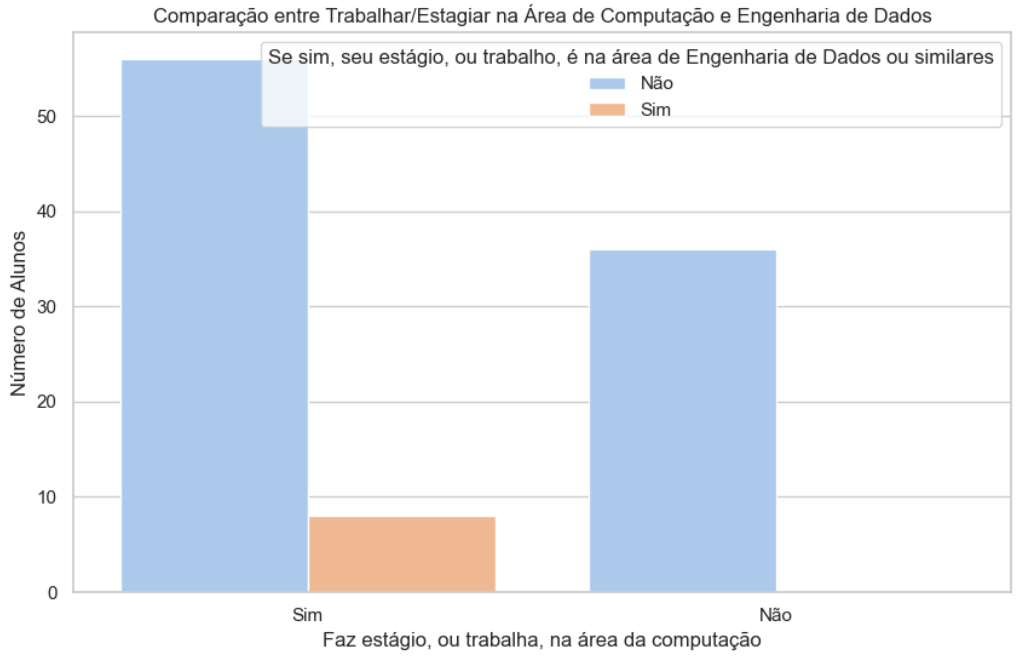


Figura 5 – Distribuição dos alunos que trabalham na área de computação.

> "As explicações teóricas e aplicações práticas foram essenciais para compreender os conceitos e conseguir aplicá-los depois, nos trabalhos, e o suporte oferecido também sempre conseguiu ajudar."

> "Me mostrou um pouco do outro da computação e o novo jeito de trabalhar com dados massivos de uma forma eficaz. Melhorei bastante quanto o conhecimento sobre manipulação e técnicas de grande volume de dados e pretendo seguir aprendendo sobre a área para logo me inserir no mercado com esse novo conhecimento."

> "Aprendi muitos conceitos importantes na área de processamento de dados e big data. Aprendi os casos de uso para bancos de dados SQL e NoSQL, diferentes algoritmos usados em clusters, processamento de dados com Spark e os processos envolvendo essa área, etc."

A maioria dos estudantes relatou uma compreensão satisfatória dos conceitos abordados, especialmente no que tange às diferenças entre plataformas de dados (Data Warehouse, Data Lake, e Lakehouse) e ao uso de bancos de dados relacionais e NoSQL.

No entanto, uma minoria expressou dificuldades em acompanhar o ritmo das aulas, sugerindo que, em alguns casos, o conteúdo foi apresentado de forma acelerada, sem tempo

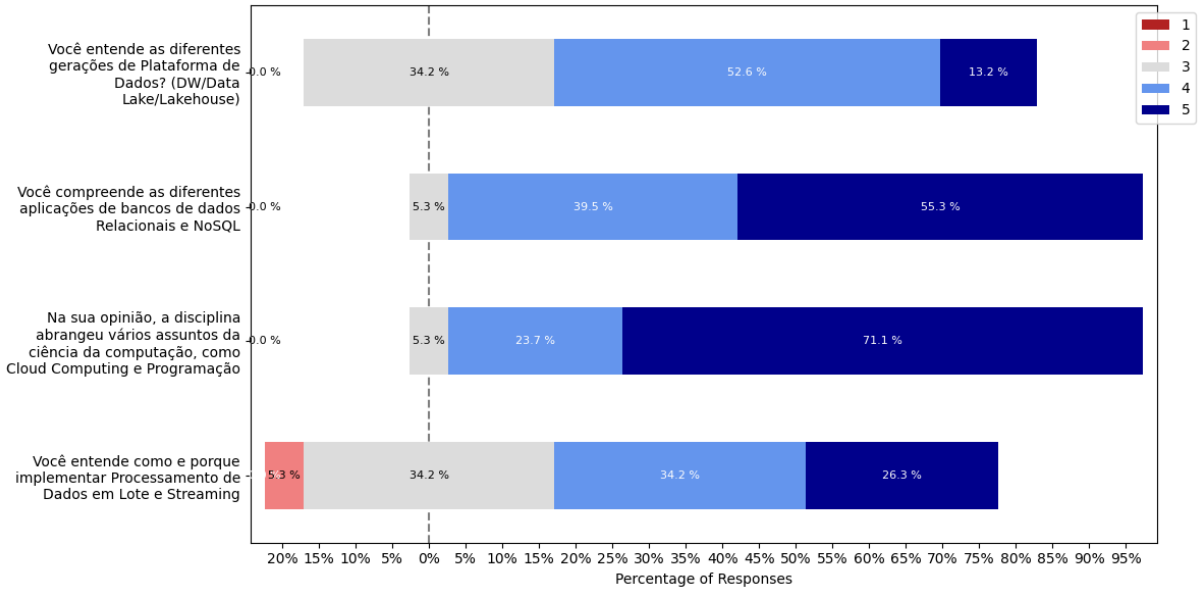


Figura 6 – Avaliação do Conteúdo Aplicado

suficiente para a assimilação.

A segunda parte do formulário avaliou a disciplina, sua dificuldade, aplicabilidade e qualidade do ensino. Os resultados estão na Figura 7

Na pergunta *A disciplina foi apresentada, com objetivos claros?*, 55,3% concordaram totalmente; 31,6% concordam; e 10,5% estão neutros, o restante discorda parcialmente. Quando questionados sobre a relevância da disciplina na sua formação, 65,8% concordam totalmente, 15,8% concordam parcialmente, 7,9% estão neutros, 7,9% discordam parcialmente e 2,6% não concordam. Avaliando o material didático disponibilizado, 36,8% avaliaram como excelente; 50% avaliaram como bom; 7,9% foram neutros e; 2,6% avaliaram como ruim e péssimo. Quanto ao nível de dificuldade da disciplina, 34,2% avaliaram como muito difícil; 44,7% como difícil e; 21,1% avaliaram como médio. Questionados sobre se os créditos disponibilizados para a disciplina serem suficientes, 42,1% concordaram totalmente; 31,6% concordam parcialmente; 23,7% foram neutros; 13,2% discordaram parcialmente e 2,6% discordaram totalmente. Sobre o espaço para prática dos alunos em laboratório durante o período, 36,8% concordam totalmente; 28,9% concordam parcialmente; 21,1% foram neutros; 7,9% discordaram parcialmente e; 5,3% discordaram totalmente. Na afirmação *Os trabalhos práticos conseguiram refletir o conteúdo ensinado*, 44,7% concordam totalmente; 31,6% concordam parcialmente; 18,4% ficaram neutros e; 5,3% discordaram parcialmente. Quando questionados sobre os pré-requisitos da disciplina estarem adequados ao conteúdo apresentado, 42,1% concordam totalmente; 26,3% concordam parcial-

mente; 21,1% ficaram neutros; 7,9% discordaram parcialmente e; 2,6% discordaram totalmente. Na afirmação *Após cursar a disciplina, seu interesse pelo assunto aumentou*, 23,7% concordam totalmente; 31,6% concordam parcialmente; 28,9% ficaram neutros; 5,3% discordaram parcialmente e; 10,5% discordaram totalmente. Por fim, quando questionados se adquiriram conhecimentos novos na disciplina, 73,7% concordam totalmente e; 26,3% concordam parcialmente.

Foi feita uma pergunta discursiva, para os alunos opinarem livremente sobre o conteúdo. Destacam-se alguns comentários:

> "A disciplina foi bem ministrada, e o assunto é complexo, porém compreensível e aplicável com um certo grau de esforço. Acredito que o aproveitamento dela será melhor quando o mesmo prof. ministrar a disciplina que é pré-requisito desta, então alunos que as pegarem em sequência vão ter mais familiaridade com a didática do prof. e o assunto deverá combinar mais entre as duas"

> "As provas e trabalhos cobrem os assuntos abordados nas aulas, porém ferramentas como o Spark tem uma documentação complexa de ser seguida. Além disso, os assuntos são extensos e acredito que devam ser subdivididos em mais provas para um melhor aproveitamento do aluno. Contudo, aprendi muitas coisas na disciplina. Apesar de não me interessar muito pela área, acredito que foi importante conhecer esses conceitos para aplicá-los no meu dia a dia quando necessário"

> "Senti dificuldade com o modelo híbrido e acho que faltou um foco para a parte prática, principalmente para quem nunca mexeu com essas ferramentas."

> "Os assuntos são bem apresentados e pertinentes. Os trabalhos são interessantes e oferecem uma boa oportunidade de prática dos conceitos apresentados nas aulas. Contudo, acredito que a aplicação de exercícios práticos menores e não avaliativos melhoraria a execução dos trabalhos avaliativos"

A maioria dos alunos dos respondentes considerou a disciplina extremamente relevante para sua formação profissional, especialmente aqueles que já trabalham ou pretendem trabalhar em áreas relacionadas ao processamento de grandes volumes de dados. Por outro lado, alguns alunos, que não atuam diretamente com grandes volumes de dados ou que trabalham em áreas distantes da engenharia de dados, mencionaram que os conceitos aprendidos são de difícil aplicação imediata. Isso sugere que, embora o conteúdo seja relevante, ele pode não se alinhar diretamente com as expectativas profissionais de todos os alunos. Muitos consideraram o tempo

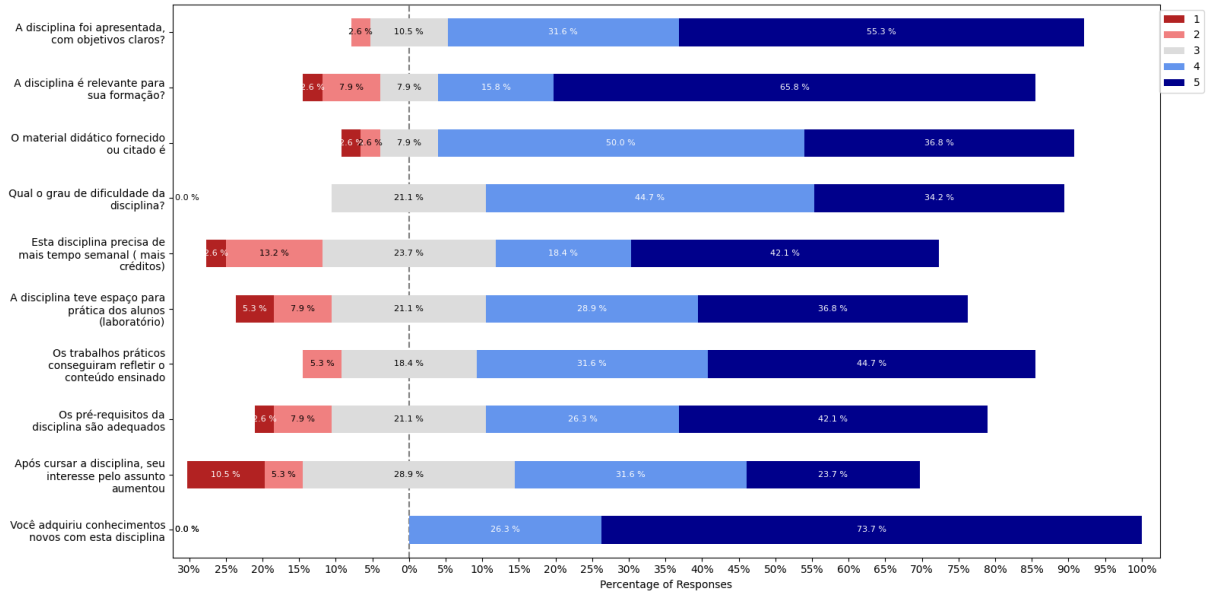


Figura 7 – Avaliação da Disciplina

disponibilizado para a disciplina insuficiente, assim como a necessidade de exercícios mais fracionados para prática deliberada.

No final do questionário, tivemos duas perguntas para que os alunos, que trabalham em alguma área da Ciência da Computação, respondessem sobre a relevância e aplicabilidade no dia a dia. Foram 15 respostas, e os resultados podem ser vistos na Figura 8. Na afirmação *Você entende serem relevantes para o trabalho do dia a dia os conteúdos ensinados*, 33,3% concordaram totalmente; 53,3% concordam parcialmente e; 13,3% foram neutros. Já na afirmação *Você entende que os conteúdos dados podem ser aplicados no dia a dia, dada sua relevância para a indústria*, 60% concordaram totalmente; 26,7% concordam parcialmente e; 13,3% foram neutros. Foi feita uma pergunta discursiva, para os alunos darem sua opinião quanto à relevância e aplicabilidade. Destacam-se alguns comentários:

> "Trabalho em um órgão do governo na área de processamento de dados. Nesse local, minha função principal é a de desenvolvedor backend, então devo conhecer as diversas tecnologias que abrangem essa área. Já tive a oportunidade de realizar processamento de uma grande quantidade de dados e apliquei alguns conceitos vistos na disciplina para aproveitar melhor a tarefa. Ainda não tive mais oportunidades, mas, se houver, estarei preparado para aplicar os conteúdos passados."

> "Atualmente eu faço parte do DX LG Academy. Uma academia de desenvolvimento de RPAs para a indústria, onde uma das vertentes é possibilitar ao membro o apren-

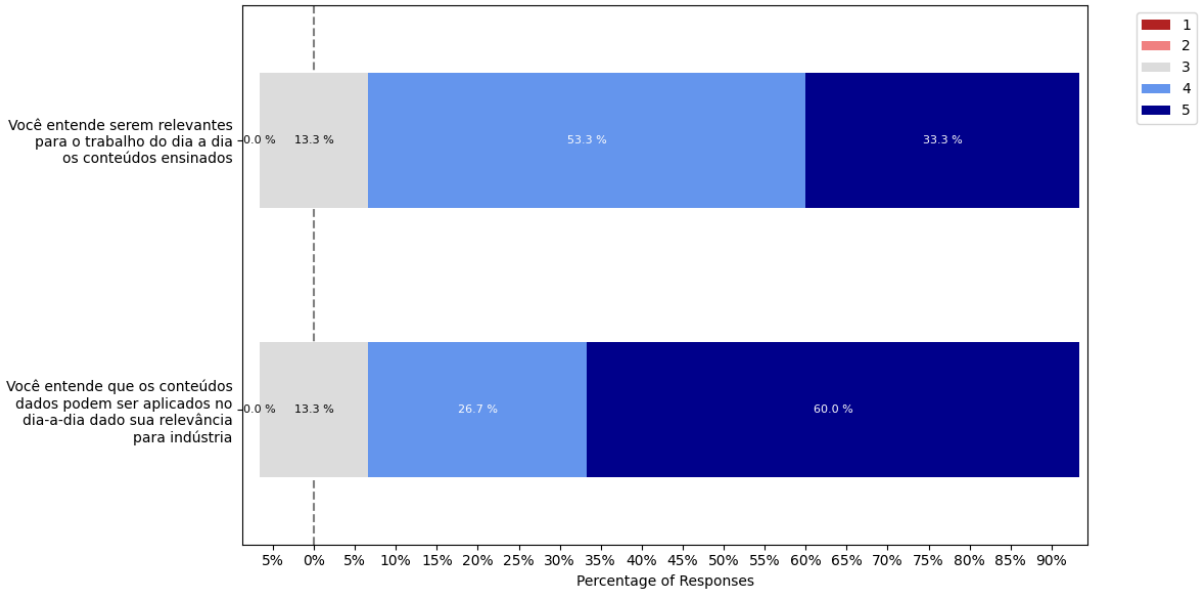


Figura 8 – Avaliação da relevância e aplicabilidade

dizado e aplicação dos conhecimentos da engenharia de dados nos problemas da indústria. Entender como os processos ETL funcionam, quais as vantagens e desvantagens da inserção de dados em batch, ou mesmo as características da computação em nuvem para o cenário da indústria faz toda a diferença na experiência da capacitação, uma vez que além da capacitação também estamos aplicando esse conhecimento em problemas reais."

6.3.2 Discussão dos resultados

A avaliação dos conteúdos indicou que a maioria dos alunos desenvolveu uma compreensão sólida dos conceitos fundamentais abordados, como a diferença entre Data Warehouse, Data Lake e Lakehouse. Além disso, a alta taxa de compreensão quanto às diferenças entre bancos de dados relacionais e NoSQL destaca a eficácia do curso em cobrir os fundamentos essenciais da disciplina. Estes tópicos são centrais para a Engenharia de Dados e refletem o sucesso da disciplina em introduzir os alunos a conceitos essenciais.

No entanto, as respostas indicam uma dificuldade relativa no entendimento do processamento em lote e *streaming*. Essa variação sugere que, embora o conteúdo tenha sido apresentado de forma abrangente, a complexidade dos conceitos de processamento em tempo real e o entendimento de casos práticos requerem mais atenção. Essa área poderia ser fortalecida com

exercícios práticos adicionais ou maior detalhamento nas aulas.

A percepção dos alunos quanto à clareza dos objetivos da disciplina foi positiva. Essa resposta reflete uma condução clara do curso, que é um fator crítico para disciplinas técnicas complexas como Engenharia de Dados. Além disso, para a maioria deles, os tópicos apresentados se alinham bem com as necessidades de desenvolvimento de habilidades na área de dados.

A análise dos comentários sugere, porém, que a intensidade e o volume do conteúdo foram desafiadores para alguns alunos, especialmente para aqueles sem uma base prática sólida em ferramentas específicas como *Spark* ou *Docker*. A dificuldade em seguir o ritmo do curso foi mencionada, com algumas sugestões para subdividir as avaliações e os trabalhos práticos, permitindo uma melhor assimilação do conteúdo ao longo do período.

Os alunos que já atuam em áreas relacionadas ao processamento de dados ou ao desenvolvimento de soluções em computação destacaram a relevância e aplicabilidade dos conhecimentos adquiridos. Os alunos veem os conteúdos como aplicáveis ao ambiente de trabalho, especialmente nas áreas de ETL, computação em nuvem e processamento de dados em batch. Esses tópicos são particularmente importantes para o setor, ao refletirem as habilidades exigidas para lidar com grandes volumes de dados em diferentes contextos de processamento e armazenamento.

No entanto, a aplicabilidade dos conceitos aprendidos parece mais limitada para alunos cujas funções estão distantes do núcleo da Engenharia de Dados. Essa observação sugere a necessidade de uma abordagem mais personalizada para os alunos que possuem uma experiência prática menor, talvez por meio de conteúdos introdutórios ou exercícios básicos voltados à prática com essas ferramentas, permitindo-lhes construir uma base sólida antes de enfrentar temas mais complexos.

Em relação à estrutura da disciplina, os alunos indicaram que os créditos e o tempo disponível foram adequados para a complexidade do conteúdo. No entanto, uma parcela dos respondentes destacou a importância de exercícios práticos menores e de natureza não avaliativa, o que proporcionaria um ambiente seguro para o desenvolvimento de habilidades práticas antes da realização de avaliações formais. Isso pode melhorar o desempenho dos alunos e reduzir a ansiedade com avaliações que exigem um conhecimento técnico mais aprofundado.

De forma geral, os resultados refletem uma avaliação positiva da disciplina, com reconhecimento da relevância dos temas para a formação profissional e uma compreensão clara dos conceitos fundamentais. Os dados indicam que o curso atende às expectativas de alunos que

desejam ingressar ou se aprimorar na área de Engenharia de Dados. No entanto, para maximizar a experiência de aprendizagem, recomenda-se uma abordagem mais gradual e prática em temas avançados, promovendo uma melhor integração entre a teoria e a prática.

6.4 Avaliação da Indústria

Foi enviado um formulário para pessoas de cargos de liderança da indústria e da sociedade civil com interesse na área de Dados como um todo, que pode ser verificado no Apêndice L, e tivemos 6 respostas no total. Esse formulário está descrito no Capítulo 2. Os respondentes tiveram acesso à ementa, ao plano de aula, à descrição dos trabalhos e às avaliações feitas durante o período letivo. A primeira afirmação foi *Os assuntos tratados são relevantes no dia a dia da indústria e da sociedade*, dentre as respostas, 50% concordaram fortemente com o conteúdo apresentado na disciplina; 33.3% concordam e; 16.7% não concordaram com o conteúdo apresentado. Na segunda afirmação, *Com essas habilidades adquiridas, os alunos conseguiriam performar bem na sua empresa*, 33.3% dos respondentes concordaram fortemente; 16.7% concordam; 33.3% foram neutros e; 16.7% não concordam. A terceira afirmação foi: *Essa disciplina consegue preparar um aluno para entrar no mercado de trabalho como Engenheiro de Dados*, nessa, 50% concordaram fortemente; 16.7% concordam; 16.7% foram neutros e; 16.7% não concordam.

Também tivemos duas perguntas discursivas. Na primeira, os respondentes poderiam sugerir assuntos relevantes que não foram abordados na disciplina. Algumas sugestões foram tratar de Pipelines de dados para algoritmos de aprendizagem de máquina, mencionado duas vezes, Evolução de Esquema, Monitoramento e Observabilidade de Pipelines, Privacidade, Segurança, Qualidade de Dados, Amazon Web Services(AWS), modelagem de dados em Data Warehouses, Linhagem de Dados, Data Mesh, e abstração para camada de metadados.

Na segunda pergunta, *Comentários gerais que julgar pertinentes.*, deixamos aberto para que respondessem livremente. Destacamos abaixo alguns comentários feitos.

> "Assuntos totalmente relevantes para a atuação no mercado"

> "No geral, o curso cobre tópicos bastante relevantes para o dia a dia e serão bastante valorizados no mercado de trabalho. Excelente trabalho!"

> "No geral, a disciplina está bem completa, abordou temas importantes na parte de engenharia de dados. Os assuntos que listei acima, acredito serem bem relevantes

pra enriquecer o conteúdo. No geral, parabéns pelo trabalho e por dar importância na adequação da grade do curso versus o mercado de trabalho."

6.4.1 Discussão dos resultados

Os dados coletados do questionário enviado a profissionais de liderança na indústria e membros da sociedade civil com interesse na área de dados oferecem uma visão externa sobre a adequação e relevância da disciplina de Engenharia de Dados. Com cinco respostas, a amostra, ainda que limitada, oferece informações relevantes sobre as expectativas do mercado de trabalho e da sociedade em relação à formação de engenheiros de dados.

Na primeira afirmação, a maioria dos respondentes indicou que considera os conteúdos tratados na disciplina como relevantes para o cotidiano da indústria e da sociedade. Isso demonstra que o curso está alinhado com as expectativas gerais dos setores de tecnologia e dados, o que é essencial para garantir que os alunos estejam em contato com temas atuais e aplicáveis. Entretanto, a presença de uma resposta discordante sugere que existe margem para diversificação dos tópicos, de modo a abranger aspectos que, eventualmente, poderiam ser mais aplicáveis a contextos específicos da indústria ou de interesse social.

Em relação à segunda afirmação, que aborda a aplicabilidade das habilidades adquiridas pelos alunos, na prática, empresarial, observa-se uma leve dispersão nas respostas. Metade dos respondentes concorda que os alunos saem capacitados para atuar em suas empresas. Esses dados sugerem que, embora a disciplina forneça uma base relevante, poderia ser aprimorada com o acréscimo de habilidades práticas mais específicas, talvez por meio de estudos de caso ou simulações de problemas reais que profissionais enfrentam em ambientes corporativos.

A terceira afirmação buscou verificar se a disciplina prepara os alunos para o mercado como engenheiros de dados. A maioria dos respondentes concordou com essa afirmação, validando a estrutura do curso como um treinamento adequado para futuras oportunidades de emprego. A presença de uma resposta neutra e outra discordante apontam, porém, para uma possível lacuna em habilidades específicas ou especializações que poderiam tornar os alunos ainda mais preparados para ingressarem na área de Engenharia de Dados. Esse dado reforça a importância de adaptar a disciplina para cobrir um leque mais amplo de habilidades e atender a uma diversidade de expectativas do mercado.

As respostas às perguntas discursivas forneceram sugestões ricas e detalhadas para

aprimorar o conteúdo da disciplina. A inclusão de tópicos como Pipelines de Dados para Aprendizagem de Máquina (mencionado duas vezes), Evolução de Esquema, Monitoramento e Observabilidade de Pipelines, Privacidade e Segurança dos Dados, Qualidade de Dados e Linhagem de Dados reflete a demanda por uma formação que aborde não apenas a engenharia de dados tradicional, mas também habilidades críticas para a gestão de dados em ambientes empresariais modernos.

Outros tópicos sugeridos, como Data Mesh e abstrações para camada de metadados, indicam o valor de explorar arquiteturas e modelos emergentes no setor de dados, o que poderia posicionar os alunos à frente das tendências do mercado. Além disso, a inclusão de uma discussão sobre ferramentas específicas, como Amazon Web Services (AWS), permitiria que os alunos obtivessem uma visão prática e contextualizada da infraestrutura em nuvem. Modelagem de Dados em Data Warehouses também foi sugerida, reforçando a importância de abordar técnicas de organização e estruturação de dados que suportem as operações de análise de negócios em grande escala.

Nos comentários gerais, a maioria dos respondentes elogiou a estrutura do curso, destacando que os temas abordados são relevantes para o mercado de trabalho. Comentários como "Assuntos totalmente relevantes para a atuação no mercado" e "o curso cobre tópicos bastante valorizados no mercado de trabalho" apontam que, no geral, a disciplina cumpre seu objetivo de formar profissionais com habilidades demandadas pelo setor. Essas repostas positivas confirmam que a estrutura central do curso está bem alinhada com as necessidades do mercado, mas também reforça que ajustes nas áreas sugeridas poderiam tornar o curso ainda mais robusto e aplicável.

7

Conclusão e Trabalhos Futuros

Este trabalho teve como objetivo principal investigar a adequação entre o ensino de Engenharia de Dados nas universidades brasileiras e as demandas atuais da sociedade e da indústria, propondo recomendações baseadas nas lacunas identificadas. Para isso, buscamos inicialmente compreender o estado atual do ensino na área, identificar as principais necessidades da sociedade e do mercado e avaliar a existência de discrepâncias entre esses dois contextos. A partir dessa análise, foram exploradas possibilidades de melhorias curriculares e metodológicas para contribuir com a formação de profissionais alinhados às necessidades atuais, sem comprometer os objetivos acadêmicos das instituições de ensino superior.

Destacamos a relevância do tema em um mundo progressivamente impulsionado por dados, em que o desafio consiste em formar profissionais capacitados para lidar com volumes crescentes e diversificados de informações. Formulamos hipóteses que nortearam este trabalho: a existência de discrepâncias entre os currículos acadêmicos e as demandas da indústria, bem como a possibilidade de implementar atualizações no currículo para minimizar essas lacunas.

Para abordar essas questões, a pesquisa foi conduzida em etapas que incluíram o levantamento das ementas e bibliografias de disciplinas relacionadas à Engenharia de Dados em 23 universidades brasileiras, além de uma análise detalhada das demandas em um contexto nacional. Com base nesses dados, foram desenvolvidas taxonomias que permitiram identificar os tópicos mais frequentes, assim como as lacunas, tanto no ensino acadêmico quanto no mercado.

Com o extenso mapeamento feito neste trabalho, criamos as taxonomias da bibliografia e das ementas, medindo a frequência dos tópicos e da indústria. Foi feita uma comparação sistemática, identificando os tópicos mais abordados pela academia e as necessidades da indústria.

Tais resultados confirmaram a hipótese inicial de uma desconexão significativa entre a

formação acadêmica e as demandas do mercado. Enquanto os currículos universitários ainda priorizam abordagens tradicionais e consolidadas, as exigências do setor empresarial refletem a necessidade de competências alinhadas a tecnologias avançadas e práticas contemporâneas, que respondam à crescente complexidade e escala dos desafios relacionados ao gerenciamento e ao processamento de dados.

Com esses resultados, dois artigos foram publicados em conferências relevantes da área de Bancos de Dados (AZEVEDO; SILVA, 2023) e da área de Educação em Computação (AZEVEDO; SILVA, 2024).

Entre as principais contribuições deste trabalho, destacamos a proposta de uma nova disciplina, denominada "Engenharia de Dados", que incorpora conteúdos teóricos e práticos alinhados às demandas atuais da sociedade. Essa disciplina foi implementada de forma experimental na UFAM, com resultados promissores, em termos de engajamento dos alunos e de relevância percebida pelo mercado. A disciplina teve boa aceitação dos alunos e da indústria. Também criamos materiais didáticos e projetos para os alunos executarem durante o período letivo, adicionando uma prática valiosa para absorção do conhecimento, com exemplos reais para aproximar a teoria da sua aplicação final.

Também são contribuições valiosas o levantamento feito das disciplinas, a normalização das suas bibliografias e ementas, assim como as taxonomias relativas a elas. A partir desse material podemos aprofundar nas discrepâncias entre academia e indústria, propondo uma possível aproximação. A metodologia deste trabalho pode ser usada como base para investigar outras áreas da Ciência da Computação, assim como outros cursos de áreas distintas.

Apesar do desenvolvimento de um projeto real durante toda a aplicação da disciplina, foi identificada a necessidade da criação de mais material interativo para prática deliberada. Dada a complexidade dos temas abordados, os alunos se beneficiariam desse tipo de abordagem.

Concluímos que a integração de tópicos contemporâneos e metodologias de ensino ativas, como atividades práticas e projetos aplicados, é não apenas viável, mas essencial para formar profissionais adequados às demandas do século XXI. Acreditamos que as recomendações e informações apresentadas neste trabalho podem servir como base para reformas curriculares em instituições de ensino superior em todo o Brasil, contribuindo para uma formação mais conectada à realidade do mercado e ao avanço da sociedade como um todo.

Trabalhos Futuros

Os resultados deste trabalho reforçam sua relevância e abrem novas perspectivas para ampliar e consolidar os avanços propostos. Além disso, surgiram questões e oportunidades que apontam para caminhos promissores de pesquisa e desenvolvimento. Entre as possibilidades de trabalhos futuros, destacamos os seguintes.

Expandir a aplicação em outras turmas para refinar material e metodologia

A aplicação da disciplina proposta em turmas maiores e com perfis variados de alunos permitirá validar e refinar ainda mais o material didático e a abordagem metodológica. Essa expansão possibilitará compreender melhor a eficácia da disciplina em diferentes contextos acadêmicos e perfis de aprendizado.

Desenvolver material didático adicional com foco em exercícios práticos de laboratório

O desenvolvimento de um conjunto mais robusto de materiais didáticos, com maior ênfase em exercícios práticos para laboratório, pode facilitar a aprendizagem dos alunos e aumentar sua confiança ao lidar com ferramentas e técnicas de Engenharia de Dados em cenários reais. Essa abordagem prática também pode contribuir para maior engajamento e retenção de conhecimento por parte dos alunos.

Adaptar a metodologia para outras áreas emergentes da Ciência da Computação

A metodologia proposta neste trabalho pode ser adaptada para outras áreas emergentes da Ciência da Computação, como Ciência de Dados, Computação em Nuvem ou Segurança. Essa adaptação poderá promover uma modernização curricular mais ampla, beneficiando diferentes especializações na área de Computação.

Realizar uma análise internacional das lacunas entre indústria e academia

Expandir o escopo desta pesquisa para uma análise internacional permitirá identificar como outras universidades e mercados ao redor do mundo enfrentam a mesma questão. Essa abordagem comparativa pode gerar conhecimento valioso para aprimorar o ensino de Engenharia de Dados no Brasil e alinhar práticas internacionais.

Atualizar periodicamente o conteúdo da disciplina para acompanhar a evolução tecnológica e os impactos ético-sociais

Considerando a rápida evolução da tecnologia, será essencial revisar periodicamente o conteúdo da disciplina proposta neste trabalho, atualizando-o com base nas novas demandas e nas tendências emergentes da Engenharia de Dados. Essa prática garantirá que a disciplina continue relevante e efetiva diante de mudanças constantes no cenário tecnológico.

Essas iniciativas complementares poderão não apenas consolidar os avanços propostos neste trabalho, mas também abrir novos caminhos para uma formação mais alinhada às necessidades contemporâneas da sociedade e da indústria tecnológica.

REFERÊNCIAS

ACM. *Data Science Curricula 2021*. 2021. Disponível em: <https://www.acm.org/binaries/content/assets/education/curricula-recommendations/dstf_ccdsc2021.pdf>. 27

ACM and IEEE. *ACM Computing Curricula 2020*. 2020. Disponível em: <https://www.acm.org/binaries/content/assets/education/curricula-recommendations/cc2020.pdf>. 27

ACM, I.; AAAI. *Computer Science Curricula 2023*. 2024. Disponível em: <https://dl.acm.org/doi/pdf/10.1145/3664191>. 22, 25, 27, 38

ARMBRUST, M. e. a. Delta lake: High-performance acid table storage over cloud object stores. In: *Proceedings of the VLDB Endowment International Conference on Very Large Data Bases*. [S.l.: s.n.], 2020. 53

ARRIETA-IBARRA, I. et al. Should we treat data as labor? moving beyond "free". *AEA Papers and Proceedings*, v. 108, p. 38–42, 2018. 11

AZEVEDO, T.; SILVA, A. Um estudo sobre ensino de engenharia de dados nas universidades brasileiras: Estado atual e perspectivas de mercado. In: *Anais do IV Simpósio Brasileiro de Educação em Computação*. Porto Alegre, RS, Brasil: SBC, 2024. p. 375–383. Disponível em: <https://sol.sbc.org.br/index.php/educomp/article/view/28206>. 17, 69

AZEVEDO, T.; SILVA, A. da. Ensino de engenharia de dados nas universidades brasileiras: Estado atual e perspectivas de mercado. In: *Anais do XXXVIII Simpósio Brasileiro de Bancos de Dados*. SBC, 2023. p. 432–438. ISSN 2763-8979. Disponível em: <https://sol.sbc.org.br/index.php/sbbd/article/view/25556>. 17, 69

BARROSO LUIZ ANDRé; HOLZLE, U. *The Datacenter as a Computer: An Introduction to the Design of Warehouse-Scale Machines*. [S.l.]: Morgan Claypool Publishers, 2013. 48

BIE, T. D. et al. Automating data science. *Commun. ACM*, v. 65, n. 3, p. 76–87, 2022. 12, 19, 22, 38, 40

BOISVERT, C. Teaching relational database fundamentals: A lack-of-progress report. In: *Proceedings of the 3rd Conference on Computing Education Practice*. New York, NY, USA: Association for Computing Machinery, 2019. (CEP '19). 30

BREWER, E. Pushing the cap: Strategies for consistency and availability. In: *IEEE Computer 45(2): 23-29*. [S.l.: s.n.], 2012. 50

BURKE, Q.; BAILEY, C. S. Camp or college? In: *Proceedings of the 50th ACM Technical Symposium on Computer Science Education, SIGCSE*. New York, NY, USA: [s.n.], 2019. p. 345–350. 29

CATTELL, R. Scalable sql and nosql data stores. In: *SIGMOD Record*. [S.l.: s.n.], 2010. 50

CHACON. *Pro Git*. [S.l.]: Springer, 2010. 49

CHAMBERS BILL; ZAHARIA, M. *Spark: The Definitive Guide: Big Data Processing Made Simple.* [S.l.]: O'Reilly Media, 2018. 51, 52

CORBELLINI MATEOS, Z. G. . S. Persisting big-data: The nosql landscape. In: *Information Systems*. [S.l.: s.n.], 2017. 50

COSTA R.; HODUN, D. *Google cloud cookbook: Practical solutions for building and deploying cloud services*. [S.l.]: O'Reilly Media, 2021. 48, 49

DAVOUDIAN, C. . L. A survey on nosql stores. In: *ACM CSUR*. [S.l.: s.n.], 2018. 50

DEAN, J.; GHEMAWAT, S. Mapreduce: Simplified data processing on large clusters. In: *OSDI '04: 6th Symposium on Operating Systems Design and Implementation.* [S.l.: s.n.], 2004. 48

DENNING, P. J. Great principles of computing. *Commun. ACM*, v. 46, n. 11, p. 15–20, 2003. 25, 26

DOUGLAS, J. D. A. Evaluating teaching quality. *Quality in Higher Education*, 2006. 31

GRILLENBERGER, A.; ROMEIKE, R. Big data - challenges for computer science education. In: *Proceedingso of the 7th Int. Conf. on Informatics in Schools: Situation, Evolution, and Perspectives, ISSEP*. [S.l.: s.n.], 2014. p. 29–40. 40

GRILLENBERGER, A.; ROMEIKE, R. Key concepts of data management – an empirical approach. In: *Proceedings of the 17th Koli Calling International Conference on Computing Education Research*. [S.l.: s.n.], 2017. p. 30–39. 22, 25, 26, 27

GUOLLA, M. Assessing the teaching quality to student satisfaction relationship: Applied customer satisfaction research in the classroom. *Journal of Marketing Theory and Practice*, 1999. 31

HARENSLAK B. P.; DE RUITER, J. R. *Data pipelines with Apache airflow*. [S.l.]: Manning, 2021. 53, 54

HASSAN, I. B.; LIU, J. Embedding data science into computer science education. In: *IEEE International Conference on Electro Information Technology EIT*. [S.l.: s.n.], 2019. p. 367–372. 25, 28, 41

HICKEY, T. J.; SALAS, P. The entrepreneur's bootcamp: A new model for teaching web/mobile development and software entrepreneurship. In: *Proceeding of the 44th ACM Technical Symposium on Computer Science Education, SIGCSE*. New York, NY, USA: [s.n.], 2013. p. 549–554. 30

IVES, Z. et al. Vldb panel summary: "the future of data(base) education: Is the cow book dead?". *SIGMOD Rec.*, v. 50, n. 3, p. 23–26, 2021. 26

KIMBALL R.; ROSS, M. *The data warehouse toolkit: The definitive guide to dimensional modeling. 3. ed*. [S.l.]: John Wiley Sons, 2013. 53

KLEPPMANN, M. *Designing data-intensive applications: The big ideas behind reliable, scalable, and maintainable systems.* [S.l.]: O'Reilly Media, 2017. 48, 51, 52

LEE, J. W. Big data strategies for government, society and policy-making. In: *Journal of Asian Finance Economics and Business*. [S.l.: s.n.], 2020. 11, 26

MONTEIRO, L.; FALCãO, T.; RODRIGUES, R. Uma abordagem para planejamento de aulas interdisciplinares com pensamento computacional para educação básica. In: *Anais do IV Simpósio Brasileiro de Educação em Computação*. Porto Alegre, RS, Brasil: SBC, 2024. p. 168–176. Disponível em: <https://sol.sbc.org.br/index.php/educomp/article/view/28185>. 32

NAZÁBAL, A. et al. Data engineering for data analytics: A classification of the issues, and case studies. *CoRR*, abs/2004.12929, 2020. 19

RITTER, G. . Scalable data management: Nosql data stores in research and practice. In: *ICDE*. [S.l.: s.n.], 2016. 50

SBC. *Ref. Curricular: Bacharelado em Ciência de Dados*. 2021. Disponível em: <https://www.sbc.org.br/documentos-da-sbc/send/131-curriculos-de-referencia/1402-ref-curricular-bacharelado-em-ciencia-de-dados>. 27

SHEBARO, B. Using active learning strategies in teaching introductory database courses. *J. Comput. Sci. Coll.*, Evansville, IN, USA, v. 33, n. 4, p. 28–36, 2018. 30

SILVA, Y. N. et al. Integrating big data into the computing curricula. In: *The 45th ACM Technical Symposium on Computer Science Education, SIGCSE*. [S.l.: s.n.], 2014. p. 139–144. 28, 38, 40, 41

STONEBRAKER, M.; ÇETINTEMEL, U. "one size fits all": An idea whose time has come and gone (abstract). In: *Proceedings of the 21st International Conference on Data Engineering, ICDE*. [S.l.: s.n.], 2005. p. 2–11. 40

TECHNAVIO. *Big Data Services Market Analysis North America, Europe, APAC, South America, Middle East and Africa - US, China, Germany, Canada, Japan - Size and Forecast 2024-2028*. [S.l.], 2024. 11

TU, Y.-C. et al. An experience report on a boot-camp style programming course. In: *Proceedings of the 49th ACM Technical Symposium on Computer Science Education, SIGCSE*. New York, NY, USA: [s.n.], 2018. p. 509–514. 29

VITALINO J. F. N.; CASTRO, M. A. N. *Descomplicando o Docker, 2a edição*. [S.l.]: Brasport, 2018. 49

WANG, Y.; MCCOEY, M.; BLUM, T. Teaching database for freshmen: A two-thread model. *J. Comput. Sci. Coll.*, Evansville, IN, USA, v. 35, n. 8, p. 33–43, 2020. 30

WIJAYA, A. *Data Engineering with Google Cloud Platform: A practical guide to operationalizing scalable data analytics systems on GCP*. [S.l.]: Packt Publishing, 2022. 48, 49

ZAHARIA, M. et al. Resilient distributed datasets: A {Fault-Tolerant} abstraction for {In-Memory} cluster computing. In: *9th USENIX symposium on networked systems design and implementation (NSDI 12)*. [S.l.: s.n.], 2012. p. 15–28. 51

ZAHARIA, M. et al. Lakehouse: A new generation of open platforms that unify data warehousing and advanced analytics. In: *11th Conference on Innovative Data Systems Research, CIDR 2021, Virtual Event, Online Proceedings*. [S.l.: s.n.], 2021. 11, 38, 40, 41, 53

ZBURIVSKY D.; PARTNER, L. *Designing cloud data platforms*. [S.l.]: Manning Publications, 2021. 48, 49, 51, 52, 53, 54

A

LISTA COMPLETA DAS DISCIPLINAS COM EMENTA E BIBLIOGRAFIA

UFS

Banco de Dados I

Ementa

Arquitetura de um Sistema de Banco de Dados. Modelos de Dados: Hierárquico, Redes, Relacional, Objeto Relacional e Orientado a Objetos. Modelo relacional: conceitos, restrições de integridade, álgebra relacional, cálculo relacional. SQL: DDL, DML, restrições de integridade, visões, autorização de acesso. Projeto Conceitual: Modelo Entidade- Relacionamento (MER). Mapeamento Modelo Entidade-Relacionamento – Modelo Relacional. Normalização. Tecnologias de conexão com banco de dados. Conceitos básicos de transações, triggers e stored procedures. Trabalho Prático.

Bibliografia

- Não informado

BANCO DE DADOS II

Ementa

Processamento e otimização de consultas. Armazenamento de dados. Índices e hashing. Transações. Concorrência. Recuperação. Segurança, integridade e privacidade. Triggers. Stored procedures. Conceitos básicos de banco de dados objeto-relacional e bancos de dados não relaci-

onais. Conceitos básicos de banco de dados distribuídos. Conceitos básicos de administração de sistemas de banco de dados.

Bibliografia

- Não informado

INTELIGÊNCIA ARTIFICIAL

Ementa

História da Inteligência Artificial. Busca não informada. Busca com Heurísticas. Busca Local. Satisfação de restrições. Representação do conhecimento. Ontologia. Inferência com Lógica de Primeira Ordem. Inferência com Lógica Nebulosa. Raciocínio Probabilístico.

Bibliografia

- Não informado

TÓPICOS ESPECIAIS EM INTELIGÊNCIA ARTIFICIAL

Bibliografia

- Não informado

BANCO DE DADOS DISTRIBUÍDOS

Ementa

Arquiteturas de Banco de Dados Distribuídos. Controle de Concorrência. Sistemas de Gerenciamento de Banco de Dados (SGBD) Distribuídos. Projeto de Banco de Dados Distribuídos. Fragmentação. Banco de Dados Heterogêneos e interoperabilidade. Trabalho Prático.

Bibliografia

- Não informado

TÓPICOS ESPECIAIS EM BANCO DE DADOS

Bibliografia

- Não informado

UFES

BANCO DE DADOS I

Ementa

Sistemas de gerência de bancos de dados. Projeto de banco de dados: conceitual, lógico e físico. Modelo de dados relacional. Linguagens de definição e de manipulação de dados. Normalização de projeto de banco de dados. Técnicas de armazenamento e indexação. Processamento e otimização de consultas. Gerenciamento de transação, controle de concorrência e recuperação.

Bibliografia

- SILBERSCHATZ, Abraham; KORTH, Henry F.; SUDARSHAN, S. Sistema de banco de dados. 3. ed. Rio de Janeiro: Elsevier, 1999.
- ELMASRI, Ramez; NAVATHE, Sham. Sistemas de banco de dados. 3. ed. São Paulo: Pearson, 2002.
- HEUSER, Carlos Alberto. Projeto de banco de dados. 6. ed. Porto Alegre: Bookman, 2008.
- Complementar DATE, C. J. Introdução a sistemas de bancos de dados. 8. ed. Rio de Janeiro, Campus: Elsevier, 2004. 865 p.
- Complementar GARCIA-MOLINA, H; ULLMAN, J.D.; WIDOM, J. Database systems: the complete book. 2. ed. Editora Pearson, 2009.
- Complementar ULLMAN, J.D.; WIDOM, J. A first course in databases systems. 2. ed. Editora Prentice Hall, 2002
- Complementar GARCIA-MOLINA, Hector; ULLMAN, Jeffrey D.; WIDOM, Jennifer. Implementação de sistemas de bancos de dados. 1. ed. Editora Campus, 2001.
- Complementar RAMAKRISHNAN, Raghu; GEHRKE, Johannes. Sistemas de gerenciamento de banco de dados. 3. ed. São Paulo, SP: McGraw-Hill, 2008. xxvii, 884 p.

INTELIGÊNCIA ARTIFICIAL

Ementa

Introdução: história e fundamentos da Inteligência Artificial (IA). Resolução de problemas: métodos de busca cega, busca heurística (métodos construtivos e de busca local) e metaheurísticas e computação natural (métodos evolucionários e de inteligência coletiva). Aprendizado de máquina: conceitos básicos, métodos de aprendizado supervisionado e nãosupervisionado, métodos de avaliação. Representação do conhecimento: formas de representação, ontologias, conhecimento incerto e difuso, sistemas baseados em conhecimento (sistemas especialistas). Estudo e uso de ferramentas para construção de aplicações de IA. Aplicações da IA em problemas práticos.

Bibliografia

- RUSSEL, S. J.; NORVING, P. Inteligência artificial. 2a. edição, Editora Elsevier, 2004.
- LUGER, G.F. Artificial intelligence: structures and strategies for complex problem solving. 4a. edição, Editora Bookman, 2004
- HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. The elements of statistical learning: data mining, inference, and prediction. 2a. edição, Editora Springer, 2009.
- Complementar EBERHART, R. C.; SHI, Y. Computational intelligence: concepts to implementations. 1a. edição, Editora Morgan Kaufman, 2007.
- Complementar RESENDE, S. Sistemas inteligentes. 1a. edição, Editora Manole, 2003.
- Complementar NILSSON, N.J. Artificial Intelligence: a new synthesis. 1a. edição, Editora Morgan Kaufman, 1998.
- Complementar WITTEN, I. H.; FRANK, E.; HALL, M.A. Data mining: practical machine learning tools and techniques. 3a. edição, Editora Morgan Kaufman, 2011.
- Complementar ALPAYDIN, E. Introduction to machine learning. 2a. edição, Editora MIT Press, 2010

BANCO DE DADOS II

Ementa

Linguagem SQL. Integridade e segurança em bases de dados: conceitos e comandos SQL. Comandos analíticos em SQL. Visões, gatilhos (triggers) e procedimentos armazenados (stored procedures). Acesso multiusuário em bases de dados. Ajuste fino de desempenho de bancos de dados. Tópicos avançados em bancos de dados.

Bibliografia

- SILBERSCHATZ, A.; KORTH, H.F.; SUDARSHAN, S., Sistema de banco de dados, 3a. edição, Editora Elsevier, 1999.
- ELMASRI, R.; NAVATHE, S., Sistemas de banco de dados, 3a. edição, Editora Pearson, 2002.
- HEUSER, C.A., Projeto de banco de dados, 6a. edição, Editora Bookman, 2008.
- Complementar DATE, C.J., Introdução a sistemas de banco de dados, 8a. edição, Editora Elsevier, 2004.
- Complementar GARCIA-MOLINA, H; ULLMAN, J.D.; WIDOM, J., Database systems: the complete book, 2a. edição, Editora Pearson, 2009.
- Complementar ULLMAN, J.D.; WIDOM, J., A first course in databases systems, 2a. edição, Editora Prentice Hall, 2002.
- Complementar GARCIA-MOLINA, H; ULLMAN, J.D.; WIDOM, J., Implementação de sistemas de bancos de dados, 1a. edição, Editora Campus, 2001.
- Complementar RAMAKRISHNAM, R; GEHRKE, J., Sistemas de gerenciamento de banco de dados, 3a. edição, Editora McGraw-Hill, 2008.

INTRODUÇÃO À CIÊNCIA DE DADOS

Ementa

Ciência de Dados: conceito, características, importância, áreas de conhecimento. Áreas Correlatas: Ciência de Dados vs Engenharia de Dados, Ciência de Dados vs Análise de Dados.

Ferramentas da Ciência de Dados: bancos de dados e big data, linguagens de programação, máquinas de aprendizagem e de previsão, computação de alto desempenho.

Bibliografia

- CADY, F. The data science handbook. 1. ed. Editora John Wiley, 2017.
- CIELEN, D.; MEYSMAN, A.D.B.; ALI, M. Introducing data science: big data, machine learning, and more, using Python tools. 1. ed. Editora Manning, 2016.
- IGUAL, L.; SEGUÍ, S. Introduction to data science: A Python approach to concepts, techniques and applications. 1. ed. Editora Springer, 2017.
- Complementar BATINI, C.; SCANNAPIECA, M. Data quality: concepts, methodology and techniques 1. ed. Editora Springer, 2006.
- Complementar OZDEMIR, S. Principles of data science. 1. ed. Editora Packt, 2016
- Complementar SKIENA, S.S. The data science design manual. 1. ed. Editora Springer, 2017.
- Complementar HASTIE, Trevor; TIBSHIRANI, Robert; FRIEDMAN, Jerome. The elements of statistical learning: data mining, inference, and prediction. 2nd ed. New York, N.Y.: Springer, 2009. xxii, 745 p.
- Complementar WITTEN, I. H.; FRANK, Eibe; HALL, Mark A. Data mining: practical machine learning tools and techniques. 3rd ed. Burlington, Mass.: Morgan Kaufmann, 2011. xxxiii, 629 p.

BANCO DE DADOS I - São Mateus

Ementa

Introdução a Banco de Dados. Sistemas de Banco de Dados Legados. Modelagem de Banco de Dados: Modelagem conceitual considerando os temas atuais tais como: genealogia, gênero, aspectos étnicos, raciais e culturais. Modelagem Lógica. Modelagem Física. Teoria e Normalização de Projeto de Banco de Dados. Linguagem SQL

Bibliografia

- ELMASRI, Ramez; NAVATHE, Sham. Sistemas de banco de dados. 6a ed. São Paulo: Pearson Education do Brasil. 2011.
- SILBERSCHATZ, Abraham; KORTH, Henry F.; SUDARSHAN, S. Sistema de banco de dados. 6a ed. Elsevier: Campus, 2012.
- ROB, Peter; CORONEL, Carlos. Sistemas de Banco de Dados: Projeto, Implementação e Gerenciamento. 8a ed. Cengage Learning. 2011.
- Complementar CAMPBELL, Laine; MAJORS, Charity. Database Reliability Engineering: Designing and Operating Resilient Database Systems. 1a ed. Editora: O'Reilly, 2017. 300p.
- Complementar DATE, C. J. Introdução a sistemas de bancos de dados. 8 ed. Rio de Janeiro: Campus: Elsevier, 2004. 865 p.
- Complementar CABRAL, Alex Lima; SANTANA FILHO, Ozeas Vieira. Modelagem e Banco de Dados. 2a ed. Editora: Senac SP, 2018. 160p.
- Complementar TEOREY, Tobey J. Projeto e Modelagem de Banco de Dados. 1a ed. Editora Elsevier, 2013. 328 p.
- Complementar ALVES, William Pereira. Banco de Dados. 1a ed. Editora Erika, 2014. 160 p.

INTELIGÊNCIA ARTIFICIAL - São Mateus

Ementa

Definições e fundamentos. Agentes inteligentes. Resolução de problemas. Planejamento e representação de conhecimento. Aprendizado supervisionado e não-supervisionado. Classificadores e Indutores. Métodos de avaliação. Técnicas de aprendizado de máquina. Introdução à teoria de jogos e neurociência. Estado da arte em Inteligência Artificial.

Bibliografia

- RUSSEL, S.; NORVIG, P. Inteligência Artificial. 3a ed. Elsevier, 2013
- BITTENCOURT, Guilherme. Inteligência Artificial: Ferramentas e Teorias. 3a ed. Editora da UFSC, 2006.

- HAYKIN, S. O. Neural Networks and Learning Machines. Pearson, 3rd ed. 2008
- Complementar TEIXEIRA, João de Fernandes. O que é inteligencia artificial. Brasiliense, 1990.
- Complementar LUGER, George F. Inteligência Artificial. 6a ed. Pearson, 2013.
- Complementar WINSTON, P. H. Artificial Intelligence. 3rd ed. Addison Wesley, 1992.
- Complementar RICH, Elaine; KNIGHT, Kevin. Artificial Intelligence. 2nd ed. McGraw-Hill, 1991.
- Complementar SOUZA, João Nunes de. Lógica para Ciência da Computação. 2a ed. Campus, 2008.

BANCO DE DADOS II - São Mateus

Ementa

Teoria de Processamento de Transações. Técnicas de Controle de Concorrência. Deadlock em Banco de Dados. Técnicas de Recuperação de Banco de Dados. Segurança de Banco de Dados. SGBDs: Funcionalidades e técnicas de gerenciamento do acesso, da persistência, da manipulação e a organização dos dados. Introdução as novas tecnologias de banco de dados.

Bibliografia

- CAMPBELL, Laine; MAJORS, Charity. Database Reliability Engineering: Designing and Operating Resilient Database Systems. 1a ed. Editora: O'Reilly, 2017. 300p
- ELMASRI, Ramez; NAVATHE, Sham. Sistemas de banco de dados. 6a ed. São Paulo: Pearson Education do Brasil. 2011.
- SILBERSCHATZ, Abraham; KORTH, Henry F.; SUDARSHAN, S. Sistema de banco de dados. 6a ed. Elsevier: Campus, 2012
- Complementar CORONEL, Carlos; MORRIS, Steven. Database Systems: Design, Implementation, & Management. 1a ed. Editora: Cengage Learning, 2014. 732p.
- Complementar DATE, C. J. Introdução a sistemas de bancos de dados. 8 ed. Rio de Janeiro: Campus: Elsevier, 2004. 865 p.

- Complementar BERMAN, J. J. Principles of Big Data: Preparing, Sharing, and Analyzing Complex Information. Amsterdam: Morgan Kaufmann, 2013. (eBook).
- Complementar KRISHNAN, K. Data Warehousing in the Age of Big Data. Amsterdam: Morgan Kaufmann, 2013. (eBook).
- Complementar ROB, Peter; CORONEL, Carlos. Sistemas de Banco de Dados: Projeto, Implementação e Gerenciamento. 8a ed. Cengage Learning. 2011.

TÓPICOS ESPECIAIS EM INTELIGÊNCIA ARTIFICIAL I - São Mateus

Ementa

Introdução à temas avançados relacionadas a aprendizado de máquina, sistemas especialistas, redes neurais artificiais, lógica fuzzy, etc.

Bibliografia

- RUSSEL, S.; NORVIG, P. Inteligência Artificial. 3a ed. Elsevier, 2013.
- BITTENCOURT, Guilherme. Inteligência Artificial: Ferramentas e Teorias. 3a ed. Editora da UFSC, 2006.
- HAYKIN, S. O. Neural Networks and Learning Machines. Pearson, 3rd ed. 2008.
- Complementar BISHOP, C. Pattern Recognition and Machine Learning. Springer. 2006.
- Complementar DUDA, R. O.; HART, P. E.; STORK, D. G. Pattern Classification. Editora Wiley. 2006.
- Complementar HASTIE, T.; Tibshirani, R.; FRIEDMAN, J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. 2nd ed. Springer, 2009.
- Complementar MITCHELL, T. M. Machine Learning. McGraw–Hill Science/Engineering/Math, 1997.
- Complementar GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. Pattern Recognition and Machine Learning Springer, Cambridge, MA, 2006.

TÓPICOS ESPECIAIS EM BANCO DE DADOS I - São Mateus

Ementa

Novas tecnologias para descoberta de conhecimento em bases de dados.

Bibliografia

- GOLDSCHMIDT, Ronaldo; PASSOS, Emanuel. Data Mining: Conceitos, técnicas, algoritmos, orientações e aplicações. 2a ed. Editora Campus, 2015.
- HAN,J; PEI,J; KAMBER, M. Data Mining: Concepts and Techniques. 3a ed. Editora Elsevier, 2011.
- GARCIA, Salvador; HERRRERA, Franciso; LUENGO, Julián. Data Preprocessing in Data Mining. Springer, 2015.
- Complementar GOMES, Elisabeth; BRAGA, Fabiane. Inteligência Competitiva em Tempos de Big Data. Analisando Informações e Identificando Tendências em Tempo Real. 1a ed. Alta Books. 2017 160p
- Complementar BERMAN, JJ. Principles of Big Data: Preparing, Sharing, and Analyzing Complex Information. Amsterdam: Morgan Kaufmann, 2013.
- Complementar KRISHNAN, Krish. Data Warehousing in the Age of Big Data. 1a ed. Morgan Kaufmann Publishers. 2013. 346p.
- Complementar ELMASRI, Ramez; NAVATHE, Sham. Sistemas de banco de dados. 6a ed. São Paulo: Pearson Education do Brasil, 2011. 808p.
- Complementar SILBERSCHATZ, Abraham; KORTH, Henry F.; SUDARSHAN, S. Sistema de banco de dados. 6a ed. Elsevier: Campus, 2012. 904p.

TÓPICOS ESPECIAIS EM BANCO DE DADOS II - São Mateus

Ementa

Novas tecnologias para recuperação de informação, análise de dados e apresentação do conhecimento.

Bibliografia

- GOLDSCHMIDT, Ronaldo; PASSOS, Emanuel. Data Mining: Conceitos, técnicas, algoritmos, orientações e aplicações, 2a ed. Editora Campus. 2015.
- HAN, J.; PEI, J.; KAMBER, M. Data Mining: Concepts and Techniques. 3a ed. Editora Elsevier, 2011. 744p.
- GARCIA, Salvador; HERRERA, Franciso; LUENGO, Julián. Data Preprocessing in Data Mining. Springer, 2015.
- Complementar GOMES, Elisabeth; BRAGA, Fabiane. Inteligência Competitiva em Tempos de Big Data. Analisando Informações e Identificando Tendências em Tempo Real. 1a ed. Alta Books. 2017. 160p.
- Complementar BERMAN, JJ. Principles of Big Data: Preparing, Sharing, and Analyzing Complex Information. Amsterdam: Morgan Kaufmann, 2013.
- Complementar KRISHNAN, Krish. Data Warehousing in the Age of Big Data. 1a ed. Morgan Kaufmann Publishers. 2013. 346p.
- Complementar ELMASRI, Ramez; NAVATHE, Sham. Sistemas de banco de dados. 6a ed. São Paulo: Pearson Education do Brasil, 2011. 808p.
- Complementar SILBERSCHATZ, Abraham; KORTH, Henry F.; SUDARSHAN, S. Sistema de banco de dados. 6a ed. Elsevier: Campus, 2012. 904p.

Unicamp

MC536 - Banco de Dados: Teoria e Pratica

Ementa

Arquiteturas de sistemas de gerenciamento de bancos de dados. Modelagem de dados: modelos conceituais e lógicos, incluindo o modelo relacional e normalização. Álgebra relacional. Linguagens de definição e de manipulação de dados. Otimização de consultas. Mecanismos de proteção, recuperação e segurança. Controle de concorrência. Bancos de dados não relacionais. Projeto e desenvolvimento de ferramentas e técnicas utilizadas na solução de problemas de sistemas de informação, utilizando bancos de dados. Modelagem, especificação, projeto e implementação de aplicações em sistemas de informação.

Bibliografia

- Sistemas de Bancos de Dados, Ramez Elmasri e Shamkant B. Navathe; Addison Wesley, 6a Edição em português (2011)
- Database Management Systems, Raghu Ramakrishnan and Johannes Gehrke; McGraw-Hill, 3rd edition (2003)
- Database System Concepts, Abraham Silberschatz, Henry Korth and S. Sudarshan; McGraw Hill, 6th Edition (2010)
- Readings in Database Systems, Joseph M. Hellerstein and Michael Stonebraker; MIT Press, 4th Edition (2005)
- Database Systems: The Complete Book, Hector Garcia-Molina, Jeffrey D. Ullman, and Jennifer Widom; Pearson Prentice Hall, 2nd edition (2008)
- Projeto de Banco de Dados, Carlos Alberto Heuser; Editora Sagra Luzzato, 6a Edição (2008)

ST567 - Banco de Dados I

Ementa

Fundamentos de sistemas de bancos de dados e conceitos relacionados com a estrutura conceitual e lógica dos bancos de dados. Modelos de dados e sua aplicação no desenvolvimento de projetos de bancos de dados. Introdução a Sistemas de Gerenciamento de Banco de Dados.

Bibliografia

- • ELMASRI, R.; NAVATHE, S. B., Sistemas de Banco de Dados, 6 ed., Addison Wesley, 2011.
- • SILBERSCHATZ, A.; KORTH, H. F. e SUDARSHAN, S., Sistema de Banco de Dados, 5 ed., Campus - Elsevier, 2006.
- • DATE, C. J., Introdução a Sistemas de Bancos de Dados, 8 ed., Campus - Elsevier, 2005.
- Bibliografia complementar: • ROB, P.; CORONEL, C., Sistemas de Banco de Dados, 8 ed., Cengage, 2011.

UFMG

DCC011-DIG INTRODUCAO A BANCO DE DADOS

Ementa

Memoria auxiliar; organizacao fisica e logica. Metodos de acesso.Estruturas de arquivos. Manipulacao de bancos de dados. Linguagense pacotes. Recuperacao de informacao.

Bibliografia

- Não informado

DCC212-DIG INTRODUÇÃO À CIÊNCIA DOS DADOS

Ementa

Manipulac¿a¿o de Dados na Forma de Tabelas e Matrizes. Me¿todo Cienti¿fico. Testes de Hipo¿teses e Intervalos de Confianc¿a (Atrave¿s da Normal e por Amostragem). Noc¿o¿es de Causalidade. Regressa¿o Linear. Verossimilhanc¿a. Fitting atrave¿s de Gradient Descent e Newton-Raphson . Regularizac¿a¿o. Fitting de Distribuic¿o¿es e Testes para Comparar Distribuic¿o¿es. Regressa¿o Logi¿stica. KNN. Overfitting, Underfitting e Validac¿a¿o Cruzada.

Bibliografia

- Não informado

DCC642-DIG INTRODUÇÃO À INTELIGÊNCIA ARTIFICIAL

Ementa

Introdução: IA, Agentes Inteligentes; Representação e Solução de Problemas através de Busca em Espaço de Estados; Representação do Conhecimento, Raciocínio Probabilístico, Aprendizado de Máquina, Aplicações de Inteligência Artificial

Bibliografia

- Não informado

DCC028-DIG INTELIGENCIA ARTIFICIAL

Ementa

Métodos de resolução de problemas. Representação do conhecimento usando lógica de predicados. Representações estruturadas do conhecimento. Sistemas avançados de resolução de problemas. Tópicos avançados.

Bibliografia

- Não informado

DCC189-DIG ARMAZENS DE DADOS

Ementa

CONCEITOS BASICOS DE SISTEMAS DE SUPORTE A DECISAO E DATA WAREHOU-SE DW. ARQUITETURA DE AMBIENTES DE DW. PRINCIPAIS COMPONENTES DE SOLUCOES. MODELAGEM DIMENSIONAL. ESTUDOS DE CASO. FERRAMENTAS E APLICACOES OLAP. TENDENCIAS

Bibliografia

- Não informado

DCC057-DIG MINERACAO DE DADOS

Ementa

PROCESSO DE DESCOBERTA DO CONHECIMENTO EM BANCOS DE DADOS. CONCEITOS BASICOS DE COLETA E ENGENHARIA DE DADOS. TECNICAS DE MINERACAO DE DADOS. ASPECTOS DE IMPLEMENTACAO. DOMINIOS DE APLICACAO.

Bibliografia

- Não informado

PUC-RIO

INF1012 Modelagem de Dados

Ementa

Modelo de entidade-relacionamento: conceitos básicos de entidade, relacionamento e atributo; generalização/especialização; entidades associativas. Modelo conceitual de dados utilizando a linguagem UML. Estudos de casos

Bibliografia

- HEUSER, C. A. PROJETO DE BANCO DE DADOS; PORTO ALEGRE: BOOKMAN, 2008.
- COUGO, PAULO SERGIO. MODELAGEM CONCEITUAL E PROJETO DE BANCO DE DADOS; RIO DE JANEIRO: CAMPUS, 1997.
- LARMAN , CRAIG. UTILIZANDO UML E PADRÕES: UMA INTRODUÇÃO À ANÁLISE E AO PROJETO ORIENTADOS A OBJETOS E AO DESENVOLVIMENTO ITERATIVO; PORTO ALEGRE: BOOKMAN, 2007.

INF1383 Bancos de Dados

Ementa

Introdução aos sistemas de gerência de bancos de dados. Projeto de banco de dados: conceitual, lógico e físico. Modelo conceitual de entidades e relacionamentos. Modelo de dados relacional. Dependências funcionais e normalização. Linguagens de definição e de manipulação de dados. Álgebra relacional e sql. Restrições de integridade e visões. Organização física de bancos de dados: técnicas de armazenamento e indexação.

Bibliografia

- ELMASRI, R.; NAVATHE, S. B. SISTEMAS DE BANCO DE DADOS: FUNDAMENTOS E APLICAÇÕES; SÃO PAULO: PEARSON EDUCATION, 2011.
- SILBERSCHATZ ,A.; KORTH , H.F; SUDARSHAN, S. SISTEMAS DE BANCOS DE DADOS.; RIO DE JANEIRO: CAMPUS, 2006.

- HEUSER, CARLOS ALBERTO. PROJETO DE BANCO DE DADOS; RIO DE JANEIRO: BOOKMAN, 2009.

INF1032 INTRODUCAO A CIENCIA DOS DADOS

Ementa

Introdução à Ciência dos Dados (Data Science). Coleta de dados de diferentes fontes. Preparação dos dados para análise. Exploração (numérica e visual) de dados. Métodos de modelagem e análise. Introdução a algoritmos de estimação de parâmetros e aprendizado de máquina. Apresentação de resultados utilizando dados, tabelas e gráficos.

Bibliografia

- Zumel; Nina; John Mount. Practical data science with R, 2014.; RJ: Manning, 2014.
- Gareth James; Robert Tibshirani; Trevor Hastie. An Introduction to Statistical Learning; RJ: Springer, 2013.
- NDA. Notas de Aula; NDA: NDA, 2017.
- Complementar: Kuhn; Max; Kjell Johnson. Applied predictive modeling; New York: Springer, 2013.
- Complementar: John Hopcroft; Ravindran Kannan. Foundations of Data Science; RJ: Cornell, 2013.
- Complementar: Tan; Pang-Ning; Michael Steinbach. Introduction to Data Mining; Boston: Pearson, 2005.
- Complementar: Yaser S. Abu-Mostafa; Malik Magdon-Ismail; Hsuan-Tien Lin. Learning From Data A Short Course; RJ: AML Book, 2012.
- Complementar: Cairo, Alberto. The Truthful Art: Data, Charts, and Maps for Communication; RJ: New Riders, 2016.

INF1340 BANCO DE DADOS II

Ementa

Criação de banco de dados; organização física; índices e alocação de dados; transações, concorrência e recuperação; segurança e autorização; cursores; procedimentos armazenados e gatilhos (triggers); avaliação e sintonia fina(tunning) de consultas.

Bibliografia

- ELMASRI, R. Sistemas de Banco de Dados; Upper Saddle River: Addison-Wesley, 2005.
- DATE, C. J. Introdução a Sistemas de Banco de Dados; Rio de Janeiro: Campus, 2004.

INF1341 BANCO DE DADOS III

Ementa

Bancos de dados distribuídos e cliente-servidor; requisitos de hardware e software e comunicação entre processos; bancos de dados heterogêneos; particionamento e alocação de dados e aplicações; arquiteturas em camadas.

Bibliografia

- MELO, R. N. Banco de Dados em Aplicações Cliente-Servidor; Rio de Janeiro: Infobook, 1997.
- ORFALI, R. Cliente/Servidor - Guia Essencial de Sobrevivência; Rio de Janeiro: IPBI Press, 1996.

INF1374 BANCO DE DADOS IV

Ementa

Apresentar os conceitos, fundamentos, técnicas e ferramentas de temas da área de bancos de dados que sejam atuais ou estejam despontando como temas de aplicabilidade em bancos de dados. Destacam-se, atualmente, os seguintes temas: sistemas de apoio a decisão apoiados em data warehouses. Exploração de bancos de dados usando técnicas de data mining (mineração de dados), bancos de dados orientados a objetos e relacional-objeto, bancos de dados em ambiente

web, bancos de dados semi-estruturados, bancos de dados em ambientes de computação móvel, tuning (sintonia) de bancos de dados, bancos de dados de genoma, técnicas para construção de sistemas de gerencia de bancos de dados e técnicas de armazenamento e recuperação de dados em sistemas de gerência de bancos de dados (sgbds).

Bibliografia

- SIHGH, H. S. Data Warehouse Conceitos, Tecnologias, Implementação e Gerenciamento; São Paulo: Makron Books, 2001.
- INMON, W. H. Como Construir o Data Warehouse; Rio de Janeiro: Campus, 1997.

UFAM

ICC200 - Banco de Dados 1

Ementa

Conceitos Básicos (Abordagem de bancos de dados, Sistemas gerenciadores de bancos de dados, Aplicações de bancos de dados); Armazenamento e recuperação de dados em memória secundária (Técnicas de organização de arquivos; Estruturas de indexação e métodos de acesso); Modelo de Dados Relacional (Conceitos básicos; Linguagem SQL; Restrições de integridade; Álgebra Relacional; Formas Normais); Aspectos Operacionais de Bancos de Dados (Transações; Controle de Concorrência; Recuperação de Falhas; Processamento de Consultas); Projeto de Banco de Dados (Modelo EntidadeRelacionamento; Extensões do MER; Mapeamento ER-Relacional; Diagramas UML).

Bibliografia

- GARCIA-MOLINA, H.; ULLMAN, J. D.; WIDOM, J. Database Systems: The Complete Book. 2 ed. Prentice Hall, 2008.
- RAMAKRISHNAN, R.; GEHRKE, J. Sistemas de Gerenciamentos de Bancos de Dados. 3 ed. McGraw Hill Brasil, 2008
- ELMASRI, R.; NAVATHE, S. B. Sistemas de Banco de Dados. 6 ed. Pearson do Brasil, 2011.

- Complementar: ELMASRI, R.; NAVATHE, S. B. Fundamentals of Database Systems. 7 ed. Addison Wesley,2015.
- Complementar: SILBERSCHATZ, A.; KORTH, H. F.; SUDARSHAN, S. Sistema de Banco de Dados. 6 ed. Campus,2012.
- Complementar: SHASHA, D. E.; BONNET, P. Database Tuning: Principles, Experiments, and Troubleshooting Techniques. Morgan Kaufmann, 2003
- Complementar: DATE, C. J. Introdução a Sistemas de Bancos de Dados. Campus, 2004.
- Complementar: TEOREY, T.; LIGHTSTONE, S.; NADEAU, T.; JAGADISH, H. V. Projeto e Modelagem de Bancos de Dados. 5 ed. Campus, 2013
- Complementar: FRANÇA, E.; GOYA, M.; PUGA, S. Banco de Dados: Implementation em SQL, PL SQL e Oracle 11g. Pearson, 2014.

ICC201 - Banco de Dados 2

Ementa

Modelo Relacional Avançado (Restrições de integridade; Álgebra Relacional; Formas Normais); Projeto de Banco de Dados (Modelo Entidade-Relacionamento; Extensões do MER; Mapeamento ER- Relacional; Diagramas UML); Desenvolvimento de Aplicações para Bancos de Dados (Interfaces para SQL; Procedimentos Armazenados; Aplicações Web) Modelos de Dados Avançados (Modelo Relacional Estendido, XML) Tópicos Avançados (Mineração de Dados, Armazéns de Dados, OLAP).

Bibliografia

- GARCIA-MOLINA, H.; ULLMAN, J. D.; WIDOM, J. Database Systems: The Complete Book. 2 ed. Prentice Hall, 2008.
- RAMAKRISHNAN, R.; GEHRKE, J. Sistemas de Gerenciamentos de Bancos de Dados. 3 ed. McGraw Hill Brasil, 2008
- ELMASRI, R.; NAVATHE, S. B. Sistemas de Banco de Dados. 6 ed. Pearson do Brasil, 2011.
- Complementar: SHASHA, D. E.; BONNET, P. Database Tuning: Principles, Experiments, and Troubleshooting Techniques. Morgan Kaufmann, 2003

- Complementar: DATE, C. J. Introdução a Sistemas de Bancos de Dados. Campus, 2004.
- Complementar: SILBERSCHATZ, A.; KORTH, H. F.; SUDARSHAN, S. Sistema de Banco de Dados. 6 ed. Campus,2012.
- Complementar: WILLIAMS, Hugh, LANE David, Web Database Applications with PHP and MySQL, 2nd Edition O'Reilly Media ISBN:978-0-596-00543-6.
- Complementar: BAEZA-YATES, R., RIBEIRO-NETO, B. Modern Information Retrieval:The Concepts and Technology behind Search, 2ª edition, 2010 ISBN13: 9780321416919.
- Complementar: HEUSER, Carlos Alberto. Projeto de Banco de Dados. 6a Ed., Bookman, 2008 ISBN 978-85- 7780-382-8.

ICC221 - Topicos avançados em Banco de Dados

Ementa

Estudo de temas específicos em Banco de Dados a critério do instrutor.

Bibliografia

- Livros e artigos especializados na área de pesquisa, a serem definidos pelo Instrutor. x0000x0000

ICC203 - Gerência de Dados na Web

Ementa

Estudo de tópicos de pesquisa cujo objetivo é explorar de forma efetiva os dados estruturados disponíveis na Web, incluindo sua interação com estado-da-arte de áreas como recuperação de informação, aprendizagem de máquina e mineração de dados e envolvendo várias classes de problemas tais como: Extração de Dados de Fontes Textuais; Coleta Focada de Páginas Web; Integração de Dados de Fontes Textuais da Web; Consultas na Web Considerando Características de Estrutura.

Bibliografia

- IEEE Transactions on Knowledge and Data Engineering (IEEE TKDE)
- ACM Transactions on Information Systems (ACM TOIS)
- Very Large Data Base Journal (VLDB Journal)
- World Wide Web Journal
- Complementar: World Wide Web Conference
- Complementar: International ACM SIGIR Conference on Research and Development of Information Retrieval (ACM SIGIR)
- Complementar: International Conference on Very Large Data Bases (VLDB)
- Complementar: International Conference on the Management of Data (ACM SIGMOD)
- Complementar: IEEE International Conference on Data Engineering (ICDE)

ICC220 - Topicos Especiais em Banco de Dados

Ementa

Leitura, redação, apresentação de artigos e a implementação de experimentos de artigos científicos na área de banco de dados e recuperação de informação. Os artigos tratam de assuntos tais como extração de dados da Web, coleta de dados, integração de dados, técnicas de busca e classificação de imagens e de documentos Web, mineração de dados, entre outros.

Bibliografia

- IEEE Transactions on Knowledge and Data Engineering (IEEE TKDE)
- ACM Transactions on Information Systems (ACM TOIS)
- Very Large Data Base Journal (VLDB Journal)
- World Wide Web Journal
- Complementar: World Wide Web Conference
- Complementar: International ACM SIGIR Conference on Research and Development of Information Retrieval (ACM SIGIR)

- Complementar: International Conference on Very Large Data Bases (VLDB)
- Complementar: International Conference on the Management of Data (ACM SIGMOD)
- Complementar: IEEE International Conference on Data Engineering (ICDE)

ICC205 - Intro a Banco de Dados

Ementa

Conceitos Básicos: A abordagem de banco de dados; Sistemas gerenciadores de bancos de dados; Aplicações de bancos de dados; Armazenamento e recuperação de dados em memória secundária: Técnicas de organização de arquivos; Estruturas de indexação e métodos de acesso; Modelo de Dados Relacional: Conceitos básicos; Notação; Restrições de integridade; Álgebra Relacional; Linguagem SQL; Formas Normais; Aspectos Operacionais de Bancos de Dados: Transações; Controle de Concorrência; Recuperação de Falhas; Processamento de Consultas; Projeto de Banco de Dados: Modelo EntidadeRelacionamento (MER); Extensões do MER; Mapeamento ER para Relacional, Modelos Orientados a Objetos, Mapeamento Modelos OO para Relacional; Programação com acesso a bancos de dados.

Bibliografia

- ELMASRI, R. ; NAVATHE, S. Sistemas de Banco de Dados. 4ª Edição. Addison Wesley, 2005.
- ELMASRI, R. ; NAVATHE, S. Fundamentals of Database Systems, 5th. Edition. Addison Wesley, 2006. ISBN-13: 978-0321369574.
- SILBERSCHATZ, Abraham ; KORTH, Henry F. Sistema de Banco de Dados. 5ª Edição. Editora Campus, 2006.
- Complementar: GARCIA-MOLINA Hector, ULLMAN, Jeffrey D., WIDOM, Jennifer. Database Systems: The Complete Book. 2ª ed. Prentice Hall, 2008.
- Complementar: SHASHA, D. E.; BONNET, P. Database Tuning: Principles, Experiments, and Troubleshooting Techniques. Morgan Kaufmann, 2003
- Complementar: DATE, C. J. Introdução a Sistemas de Bancos de Dados. Campus, 2004.

- Complementar: SILBERSCHATZ, Abraham; KORTH, Henry F.; SUDARSHAN, S. Sistema de Banco de Dados. Tradução da 5ª edição. Editora Campus, 2006.
- Complementar: RAMAKRISHNAN, R.; GEHRKE, J. Sistemas de Gerenciamentos de Bancos de Dados. 3 ed. McGraw Hill Brasil, 2008

ICC210 - Pratica em Banco de Dados

Ementa

Sistemas gerenciadores de bancos de dados; Armazenamento e recuperação de dados em memória secundária: Técnicas de organização de arquivos; Estruturas de indexação e métodos de acesso; Modelo de Dados Relacional: Linguagem SQL. Projeto de Banco de Dados: Modelo Entidade Relacionamento (MER); Extensões do MER; Mapeamento ER para Relacional, Modelos Orientados a Objetos, Mapeamento Modelos OO para Relacional; Programação com acesso a bancos de dados.

Bibliografia

- GARCIA-MOLINA, H.; ULLMAN, J. D.; WIDOM, J. Database Systems: The Complete Book. 2 ed. Prentice Hall, 2008.
- RAMAKRISHNAN, R.; GEHRKE, J. Sistemas de Gerenciamentos de Bancos de Dados. 3 ed. McGraw Hill Brasil, 2008
- ELMASRI, Ramez; NAVAT SHASHA, Dennis Elliott; BONNET, Philippe. Database Tuning: Principles, Experiments, and Troubleshooting Techniques. Morgan Kaufmann, 2003.
- Complementar: SHASHA, D. E.; BONNET, P. Database Tuning: Principles, Experiments, and Troubleshooting Techniques. Morgan Kaufmann, 2003
- Complementar: MILANI, André. Postgresql - Guia do Programador. 1ª Edição. Novatec. 2008. ISBN: 9788575221570
- Complementar: WILLIAMS, Hugh C.; M,seyed M. Aprendendo Mysql. 2ª Edição. Alta Books. 2009. ISBN: 9788576081470
- Complementar: SHAMKANT B. Sistemas de Banco de Dados, 6ª edição. Editora Pearson do Brasil, 2011.

UFPE

IF685- GERENCIAMENTO DADOS E INFORMACAO

Ementa

EMENTA: ALGORITMOS E ESTRUTURAS DE DADOS.

Bibliografia

- Não informado

IF695- BANCO DE DADOS AVANÇADOS

Ementa

INTRODUÇÃO. BD TEMPORAIS. BD ESPACIAIS. SISTEMAS DE INFORMAÇÕES GEOGRÁFICAS. BD MULTIMÍDIA. OUTROS BD AVANÇADOS. PROJETO

Bibliografia

- Não informado

IF694- BANCO DE DADOS DISTRIB. E MOVEIS

Ementa

ABORDAGEM TOP-DOWN; ABORDAGEM BOTTOM-UP; BD HOMOGÊNEO X HETEROGÊNEO; BD FEDERADOS X MULTI-BANCO DE DADOS; MÉTODOS; TÉCNICAS: FRAGMENTAÇÃO, REPLICAÇÃO E ALOCAÇÃO.

Bibliografia

- Não informado

IF696- INTEG. DADOS WEB E WAREHOUSE

Ementa

INTRODUÇÃO; BD TEMPORAIS; BD ESPECIAIS; SISTEMAS DE INFORMAÇÕES GEOGRÁFICAS; BD MULTIDIMENSIONAIS 12H (16H)

Bibliografia

- Não informado

IF692- PROJETO DE BANCO DE DADOS

Ementa

PROJETO E APLICAÇÕES EM BANCO DE DADOS.

Bibliografia

- Não informado

IF823 - ARMAZENAMENTO DE DADOS

Ementa

-

Bibliografia

- Não informado

IF693 - SIST.GERENC. DE BANCO DE DADOS

Ementa

ESTRUTURAS DE ARMAZENAMENTO; ARQUITETURAS DE SGBD; PROCESSAMENTO E OTIMIZAÇÃO DE CONSULTAS; PROCESSAMENTO DE TRANSAÇÕES; CONTROLE

DE CONCORRÊNCIA; TÉCNICAS DE RECUPERAÇÃO; SEGURANÇA E AUTORIZAÇÃO; PROVA; EXERCÍCIOS.

Bibliografia

- Não informado

IF976- BANCO DE DADOS

Ementa

CONCEITOS BÁSICOS SOBRE SGBD. PROJETO CONCEITUAL DE BANCO DE DADOS. MODELO RELACIONAL. PROJETO LÓGICO DE BANCO DE DADOS. FUNDAMENTOS DE SQL E PL CONECTIVIDADE COM BANCO DE DADOS. NOÇÕES DE XML E DE BANCO DE DADOS SEMI-ESTRUTURADOS. PROJETO PRÁTICO.

Bibliografia

- Não informado

IF991- ADMINISTRAÇÃO DE BANCO DE DADOS

Ementa

ADMINISTRAÇÃO DE DADOS. FUNDAMENTOS SOBRE ARMAZENAMENTO DE DADOS, INDEXAÇÃO E PROCESSAMENTO/OTIMIZAÇÃO DE CONSULTA. NOÇÕES DE PROCESSAMENTO DE TRANSAÇÕES, CONCORRÊNCIA E RECUPERAÇÃO. CONCEITOS PRÁTICOS SOBRE ADMINISTRAÇÃO DE UM SGBD

Bibliografia

- Não informado

UFPR

CI1218 - Banco de dados

Ementa

Conceitos básicos de bancos de dados e sistema gerenciadores de bancos de dados (SGBD); Modelos de dados e linguagens de consultas; Processamento de consultas e otimização; Projeto de bancos de dados; Conceitos de transações; Controle de concorrência e recuperação. Gasto energético e meio ambiente.

Bibliografia

1 Raghu Ramakrishnan e Johannes Gehrke. Sistemas de Gerenciamento de Banco de Dados. Terceira Edi- ção. Ed. McGraw-Hill, 2011.

2 Abraham Silberschatz, Henry F. Korth e S. Sudarshan. Sistema de Banco de Dados. Sexta Edição. Editora Campus, 2012.

3 Ramez Elmasri e Shamkant B. Navathe. Sistemas de Banco de Dados. Sexta Edicao. Editora Pearson Education Br, 2011.

- Complementar: [4] Jeffrey Ullman. Principles of Database and Knowledge-Base Systems, Volume 1. Second Edition. Computer Science Press, 1988.

- Complementar: [5] Serge Abiteboul, Richard Hull e Victor Vianu. Foundations of Databases. Addison-Wesley, 1995.

- Complementar: [6] Hector Garcia-Molina, Jeffrey D. Ullman, Jennifer Widom. Database Systems: The Complete Book. Ed. Pearson, 2nd Edition, 2009

- Complementar: [7] Carlos A. Heuser. Projeto de Banco de Dados - Vol.4. Ed. Bookman, Sexta Edição, 2010

- Complementar: [8] C. J. Date. Introdução a Sistemas de Bancos de Dados (Tradução da 8a Edição). Ed. Campus, 2004

CI 1087 Topicos em Banco de Dados

Ementa

Estudar os diferentes modelos de dados e as técnicas de armazenamento de dados em ambientes centralizados e distribuídos. Discutir os diferentes métodos de acesso e processamento de dados, a influência das diferentes cargas de trabalho e os diferentes segmentos de bancos de dados: Transacional, Data Warehouse, NoSQL, Arrays e MapReduce.

Bibliografia

- 1) Sistemas de Gerenciamento de Banco de Dados, Raghu Ramakrishnan, Johannes Gehrke, Ed. McGraw-Hill, 2007
- 2) M. Tamer Ozsu, Patrick Valduriez: Principles of Distributed Database Systems, Second Edition Prentice-Hall 1999
- 3) Sistemas de Banco de Dados - Fundamentos e Aplicacoes, Ramez Elmasri, Shamkant B. Navathe, Editora Pearson, Quarta Edicao, 6a
- Complementar: 1. Sistema de Banco de Dados, Abraham Silberschatz, Henry F. Korth, S. Sudarshan, Quinta Edicao, Editora Campus, 2006
- Complementar: 2. Hector Garcia-Molina, Jeffrey D. Ullman, Jennifer Widom: Database systems - the complete book (2. ed.). Pearson Education 2009
- Complementar: 3. Joseph M. Hellerstein, Michael Stonebraker, James R. Hamilton: Architecture of a Database System. Foundations and Trends in Databases, 2007.
- Complementar: 4. Franz Faerber, Alfons Kemper, Per-Åke Larson, Justin J. Levandoski, Thomas Neumann, Andrew Pavlo: Main Memory Database Systems. Foundations and Trends in Databases, 2017
- Complementar: 5. Shivnath Babu, Herodotos Herodotou: Massively Parallel Databases and MapReduce Systems. Foundations and Trends in Databases, 2013.
- Complementar: 6. Daniel Abadi, Peter A. Boncz, Stavros Harizopoulos, Stratos Idreos, Samuel Madden: The Design and Implementation of Modern Column-Oriented Database Systems. Foundations and Trends in Databases, 2013.

UFRJ

ICP489-Banco de Dados I

Ementa

ARQUITETURA DE SISTEMAS DE GERENCIAMENTO DE BANCO DE DADOS. MODELAGEM DE DADOS. BANCO DE DADOS RELACIONAIS. PROJETO DE BANCOS DE DADOS. CONTROLES OPERACIONAIS EM SISTEMAS DE BANCO DE DADOS.

Bibliografia

- -

ICP356-Organização de Dados II

Ementa

CONCEITO DE ARQUIVO. MODELAGEM FÍSICA DE ARMAZENAMENTO. ORGANIAZAÇÃO PRIMÁRIAS DE ARQUIVOS: SEQUENCIAL, DIRETO ("HASH"), INDEXADO SEQUENCIAL. ESTRUTURAS DE ÍNDICES: LISTAS INVERTIDAS, MAPAS, ÁRVORES. ORDENAÇÃO DE ARQUIVOS. COMPLEXIDADE DOS ALGORÍTMOS APRESENTADOS.

Bibliografia

- -

ICP491-Banco de Dados II

Ementa

HISTÓRICO DE BANCO DE DADOS. ARQUITETURA DE BANCO DE DADOS. BANCO DE DADOS ORIENTADO A OBJETOS E OBJETOS-RELACIONAIS. CAMADAS DE PERSISTÊNCIA. BANCO DE DADOS DISTRIBUÍDOS. BANCO DE DADOS AUTONÔMOS- BANCO DE DADOS AUTO-CONFIGURÁVEIS. GERENCIAMENTO DE PROCESSOS E SOA. CONCEITOS BÁSICOS DE MINERAÇÃO DE DADOS. GESTÃO DO CONHECIMENTO E ANÁLISE DE REDES SOCIAIS.

Bibliografia

- -

ICP490-Lab Banco de Dados

Ementa

Arquitetura de banco de Dados. Armazenamento de dados e indexação. Avaliação de operadores relacionais. Otimização de consultas. Projeto físico da banco de dados e "tuning". Segurança. Gerenciamento de transações e "triggers". Boas práticas na construção de consultas. Controle de concorrência. Recuperação de "crash". Análise de ferramentas de auxílio à gerência de BDs (modelo, análise de desempenho, etc).

Bibliografia

- -

ICP532-Mineração de Dados

Ementa

Considerações iniciais da área e aplicações.processo de descoberta do conhecimento(KDD) em base de dados.Análise Estatística de Dados.Classificação.Agrupamento.Associação.Análise de ferramentas de mineração de dados.Casos de aplicações:CRM,cenário científico,redução de inadimplência,detecção de fraude e outros.

Bibliografia

- -

EEL871-Banco de Dados

Ementa

Modelos conceitual, logico e fisico. Modelagem conceitual. Modelos logicos: relacional, hierarquico, redes e orientado a objetos. Modelo relacional: base conceitual, linguagens, normalizacao.

Modelo hierarquico: base conceitual, linguagens. Modelo redes: base conceitual, linguagem. Modelo orientado a objetos. Seguranca. Integridade. Recuperacao. Concorrencia. Banco de dados distribuido.

Bibliografia

- -

COS480-Construção de Bancos de Dados

Ementa

Sistemas de gerenciamento de bases de dados (SGBD): arquitetura, segurança, integridade, concorrência, recuperação após falha, gerenciamento de transações, métodos de acesso, otimização de consultas e aspectos de distribuição.

Bibliografia

- -

COS601-Bancos de Dados Avançados

Ementa

O curso de Bancos de Dados Não Convencionais faz uma caracterização das aplicações ditas não convencionais apresentando as principais tecnologias sendo utilizadas pelos SGBDs para atender às novas exigências destas aplicações. Dentre estas tecnologias está a utilização de conceitos orientados a objetos em SGBDs, uso da web, novos modelos de transação. São apresentados os conceitos de representação e manipulação de objetos em SGBDs. São apresntadas as tendências atuais nos novos Sistemas de Banco de Dados.

Bibliografia

- -

COS604-Dis.e Parale.em B.de Dados

Ementa

Aspectos de arquiteturas de sistemas de banco de dados distribuídos. Projeto de bases de dados distribuídas. Apresentação de técnicas envolvidas no processamento distribuído de consultas. Características da gerência de transações. Tendências atuais enfocando os sistemas de banco de dados distribuídos orientados a objetos e os servidores de dados distribuidos. Caracterísiticas de processamento paralelo em sistemas de banco de dados. Algoritmos de processamento paralelo de consultas. Fragmentação e alocação de objetos. Avaliação de desempenho. Análise dos principais sistemas de banco de dados com processamento paralelo.

Bibliografia

- -

COS624-Tóp Esp em Banco de Dados

Ementa

Os principais conceitos ligados à gerência de dados em diversos ambientes computacionais. Ênfase à aplicações científicas como, por exemplo, bioinformática e engenharia de petróleo. Sistemas de gerência de workflows científicos (SGWfC): apoio a experimentos em suas diversas etapas. Principais sistemas que apoiam a gerência de dados científicos com demonstrações de uso: base para as avaliações a serem desenvolvidas ao longo da disciplina. Um panorama da tecnologia atual no apoio a proveniência de dados e processos do experimento científico. As técnicas mais recentes em bancos de dados quanto à gerência de dados científicos, aos aspectos de distribuição de dados e processos em workflows e à combinação de dados de proveniências com dados científicos, dentre outros.

Bibliografia

- -

EEL890-Big-data

Ementa

Hadoop como software opren source para gerenciamento de cluster para big data. História do hadoop. Caracterização de problemas que se resolve com abordagem de big data (volume, variedade, velocidade, veracidade). HDFS (Hadoop File System) . Jobs em Big Data. Map reduce com arquitetura de software. Acessando dados de Hadoop usando Hive. NoSql (banco de dados para Big Data - bem diferente do banco de dados relacional tradicional, pois não se usam tabelas). Playes e fornecedores de hardware e serviços de big data. Exemplos práticos e estudos de software em Map Reduce.

Bibliografia

- Hadoop: The Definitive Guide, por Tom White, ISBN-10: 1449311520 Programming Hive, por Edward Capriolo, ISBN-10: 1449319335

USPSC

SCC0243 - Arquitetura de Sistemas Gerenciadores de Bases de Dados

Ementa

Programa: Introdução aos conceitos e à teoria do processamento de transações. Técnicas de controle de concorrência: protocolo de bloqueio em duas fases e outras técnicas existentes. Técnicas de recuperação de falhas: algoritmo undo/redo. Algoritmos para processamento e otimização de consultas. Bases de dados distribuídas: conceitos, fragmentação, replicação e técnicas de alocação de dados, tipos de sistemas, processamento de consultas e controle de concorrência e de recuperação. Ajuste fino de desempenho: conceitos, projeto físico de bases de dados, estruturas de indexação de arquivos.

Bibliografia

- - ELMASRI, R.; NAVATHE, S.B. Fundamentals of Database Systems, 4th edition. Pearson/Addison Wesley 2004, ISBN 0321122267.

- - SILBERSCHAT, A.; KORTH, H.F.; SUDARSHAN, S. - Sistemas de Banco de Dados, 3a edição. Editora Makron Books, 1999, ISBN 8534610738.
- Complementar: - BATINI, C.; CERI, S.; NAVATHE, S.B. Conceptual Database Design: an Entity-Relationship Approach. Benjamin/Cummings Pub. Co., 1992, ISBN 0805302441.
- Complementar: - DATE, C.J. Introdução a Sistemas de Bancos de Dados. Tradução da 7a edição americana Editora Campus, 2000. ISBN 8535205608.
- Complementar: - HEUSER, C.A. Projeto de Banco de Dados. Sagra Luzzatto, 2001. ISBN 8524105909.
- Complementar: - O'NEIL, P.; O'NEIL, E. Database: Principles, Programming, and Performance, 2nd edition, Morgan Kaufmann Publishers, 2001. ISBN 1558605800.
- Complementar: - RAMAKRISHNAN, R.; GEHRKE, J. Database Management Systems, 3rd edition. McGraw-Hill, 2003, ISBN 0072465638.
- Complementar: - SETZER,V.W. Banco de Dados: Conceitos, Modelos, Gerenciadores, Projeto Lógico, Projeto Físico, 3a edição. Edgard Blucher, 1989.
- Complementar: - TOEREY, T.; Lightstone; S.; Nadeau, T. Database Modeling & Design: Logical Design. 4th edition. Elsevier: Morgan Kaufmann Publishers, 2005. ISBN 0126853525.
- Complementar: - ULLMAN, J.D., WIDOM, J. A First Course in Database Systems. Prentice Hall, 1997. ISBN 0138613370.
- Complementar: - VOSSEN, G. Data Models, Database Languages and Database Management Systems. Addison-Wesley Pub., 1991. ISBN 0201416042.

SCC0240 - Bases de Dados

Ementa

Programa: Introdução a bases de dados e Sistemas Gerenciadores de Bases de Dados. Modelagem de dados utilizando o Modelo Entidade-Relacionamento. Modelo Relacional: conceitos, álgebra relacional e normalização. Introdução a SQL: comandos de definição de dados e de manipulação de dados. Introdução ao processamento de transações: conceitos e teoria. Aspectos de implementação dos sistemas gerenciadores de bases de dados.

Bibliografia

- - ELMASRI, R.; NAVATHE, S.B. Fundamentals of Database Systems, 4th edition. Pearson/Addison Wesley 2004, ISBN 0321122267.
- - SILBERSCHAT, A.; KORTH, H.F.; SUDARSHAN, S. - Sistemas de Banco de Dados, 3a edição. Editora Makron Books, 1999, ISBN 8534610738.
- Complementar: - BATINI, C.; CERI, S.; NAVATHE, S.B. Conceptual Database Design: an Entity-Relationship Approach. Benjamin/Cummings Pub. Co., 1992, ISBN 0805302441.
- Complementar: - DATE, C.J. Introdução a Sistemas de Bancos de Dados. Tradução da 7a edição americana Editora Campus, 2000. ISBN 8535205608.
- Complementar: - HEUSER, C.A. Projeto de Banco de Dados. Sagra Luzzatto, 2001. ISBN 8524105909.
- Complementar: - O'NEIL, P.; O'NEIL, E. Database: Principles, Programming, and Performance, 2nd edition, Morgan Kaufmann Publishers, 2001. ISBN 1558605800.
- Complementar: - RAMAKRISHNAN, R.; GEHRKE, J. Database Management Systems, 3rd edition. McGraw-Hill, 2003, ISBN 0072465638.
- Complementar: - SETZER,V.W. Banco de Dados: Conceitos, Modelos, Gerenciadores, Projeto Lógico, Projeto Físico, 3a edição. Edgard Blucher, 1989.
- Complementar: - TOEREY, T.; Lightstone; S.; Nadeau, T. Database Modeling & Design: Logical Design. 4th edition. Elsevier: Morgan Kaufmann Publishers, 2005. ISBN 0126853525.
- Complementar: - ULLMAN, J.D., WIDOM, J. A First Course in Database Systems. Prentice Hall, 1997. ISBN 0138613370.
- Complementar: - VOSSEN, G. Data Models, Database Languages and Database Management Systems. Addison-Wesley Pub., 1991. ISBN 0201416042.

SCC0244 - Mineração a partir de Grandes Bases de Dados

Ementa

Introdução aos conceitos do processo de descoberta de conhecimento. Técnicas de preparação de dados. Técnicas de redução de dados. Tarefas e técnicas de mineração de dados: classificação,

regressão, detecção de agrupamentos, descoberta de regras de associação, sumarização, modelagem de dependências, detecção de tendências e exceções. Conceitos e técnicas de visualização de dados. Mineração de dados visual. Exemplos de mineração de dados em domínios específicos, tais como: bioinformática, sistemas de informação geográfica, bases de dados de imagens, bases de dados de documentos textuais, bases de dados multimídia, sistemas Web, bases de dados espaço-temporais, sistemas baseados em data streams. Ferramentas de mineração de dados

Bibliografia

- - HAN, J.; KAMBER, M. Data Mining - Concepts and Techniques. Morgan Kaufmann Publishers, 2001. ISBN 1558604898.
- - HAND, D.J.; MANNILA, H.; SMYTH, P. Principles of Data Mining. Cambridge. MA: MIT Press, 2001.
- - WITTEN,I. H.; FRANK, E. Data Mining: Practical Machine Learning Tools and Techniques, 2nd edition. Morgan Kaufmann, 2005. ISBN 0120884070.
- Complementar: - FAYYAD, U.M.; PIATETSKY-SHAPIRO, G.; SMYTH, P.; UTHURUSAMY, R. Advances in Knowledge Discovery and Data Mining. Cambridge, MA: MIT Press, 1996. ISBN 0262560976.
- Complementar: - KANTARDZIC, M.; ZURADA, J. Next Generation of Data Mining Applications. Wiley-IEEE Press, 2005. ISBN 0471656054.
- Complementar: - Dunham, M. H. Data Mining Introductory and Advanced Topics. Prentice Hall/Pearson Education, 2003. ISBN 0130888923.
- Complementar: - REZENDE, S.O. Sistemas Inteligentes: Fundamentos e Aplicações. Monole, 2003.

SCC0245 - Processamento Analítico de Dados

Ementa

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. Modelagem multidimensional dos

dados: aspectos estáticos e dinâmicos. Representação do modelo de dados multidimensional em estruturas relacionais. Exemplos práticos de modelagem multidimensional dos dados. Visões materializadas: conceito, identificação de quais visões materializar, atualização e uso em consultas OLAP. Ferramentas de data warehousing.

Bibliografia

- - KIMBALL, R., ROSS, M. The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling, 2nd edition. John Wiley and Sons, Inc., 2002. ISBN 0-471-20024-7.
- Complementar: - Artigos de periódicos e congressos referentes ao processamento analítico de dados.
- Complementar: - BERSON, A., SMITH, S. Data Warehousing, Data Mining, and OLAP. McGraw-Hill, 1997. ISBN 0-07-006272-2.
- Complementar: - IMHOFF, C., GALEMMO, N., GEIGER, J.G. Mastering Data Warehouse Design: Relational and Dimensional Techniques. Wiley Publishing Inc., 2003. ISBN 0-471-32421-3.
- Complementar: - INMON, W.H. Building the Data Warehouse, 4th edition.Wiley Publishing Inc,2005.ISBN 0-7645-9944-5.
- Complementar: - KIMBALL, R., CASERTA, J. The Data Warehouse ETL Toolkit: Practical Techniques for Extracting, Cleaning, Conforming, and Delivering Data. Wiley Publishing Inc., 2004. ISBN 0-764-56757-8.
- Complementar: - KIMBALL, R., ROSS, M., THORNTHWAITE, W., MUNDY, J., BECKER, B. The Data Warehouse Lifecycle Toolkit: Practical Techniques Building Data Warehouse and Business Intelligence Systems, 2nd edition. Wiley Publishing Inc., 2008. ISBN 0-470-14977-9.

SCC0246 - Recuperação de Dados por Conteúdo

Ementa

Definição do Problema e indicação dos conceitos envolvidos. Algoritmos de extração de características de dados complexos. Tratamento de dados em domínios que não suportam relação de ordem: imagens, vídeo, áudio, séries temporais, series síncronas e transacionais, informações

genéticas, etc. Definição de consultas por similaridade: consulta por abrangência (range-query) e aos k-vizinhos mais próximos (kNN). Métodos de acesso métricos, algoritmos de busca e tratamento da maldição de alta dimensionalidade. Previsão de seletividade e custo computacional de operações de processamento e recuperação de dados. Técnicas avançadas de recuperação de dados para aplicações especiais: data mining, data warehouse, OLAP, apoio à decisão, etc.

Bibliografia

- - FALOUTSOS, C. "Searching Multimedia Databases by Content"(Advances in Database Systems), Kluwer Academic Pub, 1996.
- - KWOK, S.H., ZHAO, J.L. "Content-based object organization for efficient image retrieval in image databases", 2006.
- - MANOLOPOULOS, Y., NANOPOULOS, A., TOUSIDOU, E. "Advanced Signature Indexing for Multimedia and Web Applications"(Advances in Database Systems).
- - MARQUES, O., FURHT, B. "Content-Based Image and Video Retrieval" (Multimedia Systems and Applications), Kluwer Academic Pub, 2002.
- - SAMET, H. "Foundations of Multidimensional and Metric Data Structures", The Morgan Kaufmann Series in Computer Graphics, 2007.
- - ZEZULA, P., AMATO, G., DOHNAL, V., BATKO, M. "Similarity Search: The Metric Space Approach"(Advances in Database Systems), Kluwer Academic Pub, 2006.

SCC0275 - Introdução à Ciência de Dados

Ementa

Ciência de Dados; Descoberta de Conhecimento em Bases de Dados; Mineração de Dados, Preparação de Dados, Pré-processamento de Dados; Modelagem de dados; Estudo de algoritmo preditivo simples (k-NN), Planejamento de experimentos; Análise de resultados experimentais.

Bibliografia

- Faceli, Katti; Lorena, Ana Carolina; Gama, João ; de Carvalho, A. C. P. L. F. (2011). Inteligência Artificial - Uma Abordagem de Aprendizado de Máquina. 1. ed. Rio de Janeiro

LTC.

- Provost, F.; Fawcett, T. Data Science for Business: What you need to know about data mining and data-analytic thinking by O'Reilly Media, 2013.
- Han, J.; Kamber, M.; Pei, J. Data Mining: Concepts and Techniques, Third Edition (The Morgan Kaufmann Series in Data Management Systems). Morgan Kaufmann , 2011.
- Witten, I.; Frank, E. Third Edition (The Morgan Kaufmann Series in Data Management Systems). 2011.
- Tan, P.-N.; Steinbach, M.; Kumar, T. Introduction to Data Mining. Addison Wesley, 2005.

SSC0158 Computação em Nuvem e Arquitetura Orientadas a Serviços

Ementa

Escopo da computação em nuvem e arquiteturas orientadas a serviços: Contexto, motivação, objetivos, conceitos e recursos necessários. Definições para Computação em Nuvem: Abordagem geral de computação em nuvem e sua aplicabilidade para o desenvolvimento de aplicações distribuídas. Vertentes de computação em nuvem: software, plataforma, infraestrutura como serviços, entre outras. Definições para Arquiteturas Orientadas a Serviços x0096 SOA e Web Services: Introdução ao modelo arquitetural, instalação, configuração e desenvolvimento de aplicações orientadas a serviços. Implantação de aplicações na nuvem. Princípios Gerais e tipos de Virtualização: Introdução, características e tipos de virtualizadores, vantagens e desvantagens. Gerenciamento de Recursos na Nuvem: Desafios e soluções que envolvem o escalonamento, provisionamento e migração de recursos na nuvem. Desenvolvimento e avaliação de desempenho de aplicações distribuídas na nuvem: Ferramentas, modelos e técnicas que permitem o desenvolvimento de aplicações na nuvem. Avaliação de desempenho em um ambiente de nuvem.

Bibliografia

- -Rajkumar Buyya, James Broberg, Andrzej M. Goscinski. Cloud Computing: Principles and Paradigms, ISBN: 0470940093, 9780470940099, . Wiley, 664 pages, 2010.

- Complementar: -Hwang, K., Dongarra, J., Fox, G.C. Distributed and Cloud Computing: From Parallel Processing to the Internet of Things. ISBN: 9780128002049, Elsevier, 2013.
- Complementar: -David S. Linthicum. Cloud Computing and SOA Convergence in Your Enterprise: A Step-by-Step Guide, ISBN-10: 0136009220, Addison-Wesley Professional; 264 pages, 1 edition, 2009.
- Complementar: -Thomas Erl, Zaigham Mahmood, Ricardo Puttini. Cloud Computing: Concepts, Technology & Architecture. Prentice Hall. ISBN-10: 0133387526, 528 pages, 2013.

SCC0540 - Bases de Dados

Ementa

Introdução a banco de dados e sistemas gerenciadores de banco de dados. Arquitetura ANSI-X3-SPARC: nível físico, nível conceitual e nível externo. Modelagem de dados utilizando o modelo entidade-relacionamento. O modelo relacional: conceitos, álgebra relacional e normalização. Introdução a SQL: comandos de definição de dados e de manipulação de dados. Introdução ao processamento de transações: conceitos e teoria. Aspectos de implementação dos sistemas gerenciadores de banco de dados.

Bibliografia

- - ELMASRI, R.; NAVATHE, S.B. Fundamentals of Database Systems, 4th edition. Pearson/Addison Wesley 2004, ISBN 0321122267.
- - SILBERSCHAT, A.; KORTH, H.F.; SUDARSHAN, S. - Sistemas de Banco de Dados, 3a edição. Editora Makron Books, 1999, ISBN 8534610738.
- Complementar: - BATINI, C.; CERI, S.; NAVATHE, S.B. Conceptual Database Design: an Entity-Relationship Approach. Benjamin/Cummings Pub. Co., 1992, ISBN 0805302441.
- Complementar: - DATE, C.J. Introdução a Sistemas de Bancos de Dados. Tradução da 7a edição americana Editora Campus, 2000. ISBN 8535205608.
- Complementar: - HEUSER, C.A. Projeto de Banco de Dados. Sagra Luzzatto, 2001. ISBN 8524105909.
- Complementar: - O'NEIL, P.; O'NEIL, E. Database: Principles, Programming, and Performance, 2nd edition, Morgan Kaufmann Publishers, 2001. ISBN 1558605800.

- Complementar: - RAMAKRISHNAN, R.; GEHRKE, J. Database Management Systems, 3rd edition. McGraw-Hill, 2003, ISBN 0072465638.

- Complementar: - SETZER,V.W. Banco de Dados: Conceitos, Modelos, Gerenciadores, Projeto Lógico, Projeto Físico, 3a edição. Edgard Blucher, 1989.

- Complementar: - TOEREY, T.; Lightstone; S.; Nadeau, T. Database Modeling & Design: Logical Design. 4th edition. Elsevier: Morgan Kaufmann Publishers, 2005. ISBN 0126853525.

B

Lista de Disciplinas Selecionadas

Instituição	Disciplinas para Incluir ou Excluir
UFS	**Incluir:** • Banco de Dados I • Banco de Dados II • Banco de Dados Distribuídos • Tópicos Especiais em Banco de Dados **Excluir:** • Inteligência Artificial • Tópicos Especiais em Inteligência Artificial

Instituição	**Disciplinas para Incluir ou Excluir**
UFES	**Incluir:** • Banco de Dados I • Banco de Dados II • Banco de Dados I - São Mateus • Banco de Dados II - São Mateus • Tópicos Especiais em Banco de Dados I - São Mateus • Tópicos Especiais em Banco de Dados II - São Mateus **Excluir:** • Inteligência Artificial • Inteligência Artificial - São Mateus • Tópicos Especiais em Inteligência Artificial I - São Mateus • Introdução à Ciência de Dados
Unicamp	**Incluir:** • MC536 - Banco de Dados: Teoria e Prática • ST567 - Banco de Dados I

Instituição	**Disciplinas para Incluir ou Excluir**
UFMG	**Incluir:** • DCC011 - Introdução a Banco de Dados • DCC189 - Armazéns de Dados • DCC057 - Mineração de Dados **Excluir:** • DCC212 - Introdução à Ciência dos Dados • DCC642 - Introdução à Inteligência Artificial • DCC028 - Inteligência Artificial
PUC-RIO	**Incluir:** • INF1012 - Modelagem de Dados • INF1383 - Bancos de Dados • INF1340 - Banco de Dados II • INF1341 - Banco de Dados III • INF1374 - Banco de Dados IV **Excluir:** • INF1032 - Introdução à Ciência dos Dados

Instituição	**Disciplinas para Incluir ou Excluir**
UFAM	**Incluir:** • ICC200 - Banco de Dados 1 • ICC201 - Banco de Dados 2 • ICC221 - Tópicos Avançados em Banco de Dados • ICC205 - Introdução a Banco de Dados • ICC210 - Pratica em Banco de Dados **Excluir:** • ICC203 - Gerência de Dados na Web
UFPE	**Incluir:** • IF685 - Gerenciamento de Dados e Informação • IF695 - Banco de Dados Avançados • IF694 - Banco de Dados Distribuídos e Móveis • IF696 - Integração de Dados Web e Warehouse • IF692 - Projeto de Banco de Dados • IF823 - Armazenamento de Dados • IF693 - Sistemas Gerenciadores de Banco de Dados • IF976 - Banco de Dados • IF991 - Administração de Banco de Dados
UFPR	**Incluir:** • CI1218 - Banco de Dados • CI1087 - Tópicos em Banco de Dados

Instituição	**Disciplinas para Incluir ou Excluir**
USP/SC	**Incluir:** • SCC0243 - Arquitetura de Sistemas Gerenciadores de Bases de Dados • SCC0240 - Bases de Dados • SCC0246 - Recuperação de Dados por Conteúdo • SSC0158 - Computação em Nuvem e Arquitetura Orientadas a Serviços • SCC0540 - Bases de Dados • SSC0965 - Streaming de Dados, Microsserviços e Containers • SCC0244 - Mineração a partir de Grandes Bases de Dados **Excluir:** • SCC0245 - Processamento Analítico de Dados • SCC0275 - Introdução à Ciência de Dados

Instituição	**Disciplinas para Incluir ou Excluir**
UFRJ	**Incluir:** • ICP489 - Banco de Dados I • ICP356 - Organização de Dados II • ICP491 - Banco de Dados II • ICP490 - Lab Banco de Dados • EEL871 - Banco de Dados • COS480 - Construção de Bancos de Dados • COS601 - Bancos de Dados Avançados • COS604 - Disqq.e Parale.em B.de Dados • EEL890 - Big Data • COS624 - Tópicos Especiais em Banco de Dados • ICP532 - Mineração de Dados

Instituição	**Disciplinas para Incluir ou Excluir**
UFSC	**Incluir:** • INE5423 - Banco de Dados 1 • INE5432 - Banco de Dados 2 • INE5454 - Tópicos Especiais em Gerência de Dados • INE5457 - Big Data Management and Mining • INE5600 - Banco de Dados 3 • INE5613 - Bancos de Dados I • INE5616 - Bancos de Dados II • DEC7129 - Banco de Dados I • DEC7134 - Banco de Dados II • INE5643 - Data Warehouse • INE5644 - Data Mining
UFRGS	**Incluir:** • INF01124 - Classificação e Pesquisa de Dados • INF01145 - Fundamentos de Banco de Dados • INF01006 - Projeto de Banco de Dados • INF01014 - Sistemas de Banco de Dados Distribuídos **Excluir:** • INF01021 - Projeto de Hiperdocumentos

Instituição	**Disciplinas para Incluir ou Excluir**
UFG	**Incluir:** • INF0022 - Banco de Dados • INF0340 - Sistemas Gerenciadores de Banco de Dados • INF0355 - Armazém de Dados • INF0357 - Banco de Dados Distribuídos • INF0363 - Integração de Dados
UFMS	**Incluir:** • 18030000796 - Banco de Dados I • 18030000869 - Banco de Dados II • 18030001741 - Tópicos em Banco de Dados • 19190004280 - Banco de Dados • 19190004000 - Laboratório de Banco de Dados • 19190002434 - Tópicos em Banco de Dados • 32010001764 - Banco de Dados
UNIRIO	**Incluir:** • TIN0120 - Bancos de Dados 1 • TIN0169 - Bancos de Dados II • TIN0135 - Administração de Bancos de Dados • TIN0137 - Tópicos Avançados em Bancos de Dados I • TIN0162 - Tópicos Avançados em Bancos de Dados II • TIN0163 - Tópicos Avançados em Bancos de Dados III

Instituição	**Disciplinas para Incluir ou Excluir**
USP/SP	**Incluir:** • ACH2004 - Bancos de Dados 1 • ACH2025 - Bancos de Dados 2 • ACH2066 - Tópicos Especiais em Bancos de Dados **Excluir:** • ACH0021 - Tratamento e Análise de Dados / Informações • ACH2177 - Introdução à Ciência de Dados

C

Bibliografia das disciplinas completa e normalizada

Livro
Titulo: Sistemas de Bancos de Dados **Autores**: Ramez Elmasri, Shamkant B. Navathe **Isbn-13**: 9788588639171.0 **Isbn-10**: 8588639173 **Versão brasileira**: Sim **Edição**: 5.0
Titulo: Sistemas de Bancos de Dados **Autores**: Ramez Elmasri, Shamkant B. Navathe **Isbn-13**: 9788579360855.0 **Isbn-10**: 8579360854 **Versão brasileira**: Sim **Edição**: 6.0
Titulo: PROJETO DE BANCO DE DADOS **Autores**: CARLOS ALBERTO HEUSER **Isbn-13**: 9788577803828.0 **Isbn-10**: 8577803821 **Versão brasileira**: Sim
Titulo: Introdução a Sistemas de Bancos de Dados **Autores**: DATE, C. J. **Isbn-13**: 9788535205602.0 **Isbn-10**: 8535205608

Versão brasileira: Sim
Titulo: Introdução a Sistemas de Bancos de Dados **Autores**: DATE, C. J. **Isbn-13**: 9788535205602.0 **Isbn-10**: 8535205608 **Versão brasileira**: Sim
Titulo: Sistemas de Bancos de Dados **Autores**: Ramez Elmasri, Shamkant B. Navathe **Isbn-13**: 9788588639171.0 **Isbn-10**: 8588639173 **Versão brasileira**: Sim **Edição**: 5.0
Titulo: Sistemas de Bancos de Dados **Autores**: Ramez Elmasri, Shamkant B. Navathe **Isbn-13**: 9788588639171.0 **Isbn-10**: 8588639173 **Versão brasileira**: Sim **Edição**: 5.0
Titulo: Introdução a Sistemas de Bancos de Dados **Autores**: DATE, C. J. **Isbn-13**: 9788535212730.0 **Isbn-10**: 8535212736 **Versão brasileira**: Sim
Titulo: Bancos de Dados em Aplicações Cliente/Servidor **Autores**: NASCIMENTO, RUBENS, NASCIMENTO, SIDLEY DIAS DA, TANAKA, ASTERIO KIYOSHI **Isbn-10**: 8573310545 **Versão brasileira**: Sim
Titulo: Database Administration: The Complete Guide to Practices and Procedures **Autores**: Mullins, Craig S. **Isbn-13**: 9780201741292.0 **Isbn-10**: 201741296 **Versão brasileira**: Não
Titulo: Database Tuning: Principles, Experiments, and Troubleshooting Techniques

Autores: Shasha, Dennis; Bonnet, Philippe

Isbn-13: 9781558607538.0

Isbn-10: 1558607536

Versão brasileira: Não

Titulo: Estatística básica

Autores: Morettin,Pedro Alberto,Bussab,Wilton de Oliveira

Isbn-13: 9788502207998.0

Isbn-10: 8502207997

Versão brasileira: Sim

Edição: 8.0

Titulo: NOCOES DE PROBABILIDADE E ESTATISTICA

Autores: LIMA, ANTONIO CARLOS PEDROSO DE, MAGALHAES, MARCOS NASCIMENTO

Isbn-10: 8588697025

Versão brasileira: Sim

Edição: 7.0

Titulo: Sistemas de Bancos de Dados

Autores: Ramez Elmasri, Shamkant B. Navathe

Isbn-13: 9786550110512.0

Versão brasileira: Sim

Edição: 7.0

Titulo: Sistemas de Bancos de Dados

Autores: KORTH, H. F.; SILBERSCHATZ, A.; SUDARSHAN, S.

Isbn-13: 9788535245356.0

Isbn-10: 8535245359

Versão brasileira: Sim

Titulo: PROJETO DE BANCO DE DADOS

Autores: CARLOS ALBERTO HEUSER

Isbn-13: 9788577804528.0

Isbn-10: 8577804526

Versão brasileira: Sim

Titulo: SISTEMAS DE GERENCIAMENTO DE BANCO DE DADOS

Autores: JOHANNES GEHRKE, RAGHU RAMAKRISHNAN

Isbn-13: 9788577261314.0

Versão brasileira: Sim

Edição: 3.0

Titulo: Introdução a Sistemas de Bancos de Dados

Autores: DATE, C. J.

Isbn-13: 9788535212730.0

Isbn-10: 8535212736

Versão brasileira: Sim

Edição: 8.0

Titulo: SISTEMAS DE BANCO DE DADOS: DESIGN, IMPLEMENTAÇÃO E ADMINISTRAÇÃO

Autores: Carlos Coronel

Isbn-13: 9788522107865.0

Isbn-10: 8522107866

Versão brasileira: Não

Titulo: Sistemas de Bancos de Dados

Autores: Ramez Elmasri, Shamkant B. Navathe

Isbn-13: 9786550110512.0

Versão brasileira: Sim

Edição: 7.0

Titulo: Sistemas de Bancos de Dados

Autores: KORTH, H. F.; SILBERSCHATZ, A.; SUDARSHAN, S.

Isbn-13: 9788535245356.0

Isbn-10: 8535245359

Versão brasileira: Sim

Titulo: SISTEMAS DE GERENCIAMENTO DE BANCO DE DADOS

Autores: JOHANNES GEHRKE, RAGHU RAMAKRISHNAN

Isbn-13: 9788577261314.0

Versão brasileira: Sim

Edição: 3.0

Titulo: Introdução a Sistemas de Bancos de Dados

Autores: DATE, C. J.

Isbn-13: 9788535212730.0

Isbn-10: 8535212736

Versão brasileira: Sim

Edição: 8.0
Titulo: Database Tuning: Principles, Experiments, and Troubleshooting Techniques **Autores**: Shasha, Dennis; Bonnet, Philippe **Isbn-13**: 9781558607538.0 **Isbn-10**: 1558607536 **Versão brasileira**: Não
Titulo: -Communications of the ACM
Titulo: -Journal of the ACM
Titulo: -ACM Computing Surveys
Titulo: -ACM Transactions on Database Systems (TODS)
Titulo: -ACM Transactions on Office Information Systems (TOIS)
Titulo: -ACM Transactions on Information Systems (TOIS)-
Titulo: -ACM Transactions on Graphics (TOG)
Titulo: -SIGMOD Record
Titulo: Data Mining: Practical Machine Learning Tools and Techniques **Autores**: Witten, Ian H.; Frank, Eibe; Hall, Mark A.; Pal, Christopher J. **Isbn-13**: 9780128042915.0 **Isbn-10**: 128042915 **Versão brasileira**: Não **Edição**: 4.0
Titulo: An Introduction to Categorical Data Analysis **Autores**: Agresti, Alan **Isbn-13**: 9780471226185.0 **Isbn-10**: 471226181 **Versão brasileira**: Não **Edição**: 2.0
Titulo: Mining of Massive Datasets **Autores**: Leskovec, Jure **Isbn-13**: 9781107077232.0 **Isbn-10**: 1107077230 **Versão brasileira**: Não **Edição**: 2.0
Titulo: Data Mining and Data Warehousing: Principles and Practical Techniques **Autores**: Bhatia, Parteek

Isbn-13: 9781108727747.0

Isbn-10: 1108727743

Versão brasileira: Não

Edição: 1.0

Titulo: Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data

Autores: EMC

Isbn-13: 9781118876138.0

Isbn-10: 111887613X

Versão brasileira: Não

Titulo: Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems

Autores: Géron, Aurélien

Isbn-13: 9781491962299.0

Isbn-10: 1491962291

Versão brasileira: Não

Edição: 1.0

Titulo: Introduction to Data Science: Data Analysis and Prediction Algorithms with R

Autores: IRIZARRY, R.A

Isbn-13: 9780367357986.0

Isbn-10: 367357984

Versão brasileira: Não

Edição: 1.0

Titulo: Applied Predictive Modeling

Autores: Johnson; Kjell; Kuhn; Max

Isbn-13: 9781461468486.0

Isbn-10: 1461468485

Versão brasileira: Não

Titulo: Modelos de Regressão com Apoio Computacional

Autores: PAULA, G.A

Versão brasileira: Sim

Edição: 1.0

Titulo: Critical analysis of Big Data challenges and analytical methods.

Autores: SIVARAJAH, U., ET AL

Paper: https://doi.org/10.1016/j.jbusres.2016.08.001

Versão brasileira: Não

Titulo: Fundamentals of Database Systems

Autores: Elmasri, Ramez; Navathe, Shamkant B.

Isbn-13: 9780321122261.0

Isbn-10: 321122267

Versão brasileira: Não

Edição: 4.0

Titulo: Sistemas de Bancos de Dados

Autores: KORTH, H. F.; SILBERSCHATZ, A.; SUDARSHAN, S.

Isbn-13: 9780070310865.0

Isbn-10: 70310866

Versão brasileira: Não

Edição: 3.0

Titulo: Conceptual Database Design: an Entity-Relationship Approach

Autores: BATINI, C.; CERI, S.; NAVATHE, S.B.

Isbn-13: 9780805302448.0

Isbn-10: 805302441

Versão brasileira: Não

Edição: 1.0

Titulo: Introdução a Sistemas de Bancos de Dados

Autores: DATE, C. J.

Isbn-13: 9788535205602.0

Isbn-10: 8535205608

Versão brasileira: Sim

Edição: 7.0

Titulo: PROJETO DE BANCO DE DADOS

Autores: CARLOS ALBERTO HEUSER

Isbn-13: 9788524105906.0

Isbn-10: 8524105909

Versão brasileira: Sim

Titulo: Database: Principles, Programming, and Performance

Autores: O'NEIL, P.; O'NEIL, E.

Isbn-13: 9781558605800.0

Isbn-10: 1558605800

Versão brasileira: Não

Edição: 2.0

Titulo: SISTEMAS DE GERENCIAMENTO DE BANCO DE DADOS

Autores: JOHANNES GEHRKE, RAGHU RAMAKRISHNAN

Isbn-13: 9788577261314.0

Versão brasileira: Sim

Edição: 3.0

Titulo: Banco de Dados: Conceitos, Modelos, Gerenciadores, Projeto Lógico, Projeto Físico

Autores: SETZER,V.W.

Versão brasileira: Sim

Edição: 3.0

Titulo: Projeto e Modelagem de Banco de Dados

Autores: Teorey, Toby J.; Lightstone, Sam S.; Nadeau, Tom; Jagadish, H.V.

Isbn-13: 9780126853520.0

Isbn-10: 126853525

Versão brasileira: Não

Edição: 4.0

Titulo: First Course in Database Systems

Autores: ULLMAN, J.D., WIDOM, J.

Isbn-13: 9780138613372.0

Isbn-10: 138613370

Versão brasileira: Não

Edição: 1.0

Titulo: Data Models, Database Languages, and Database Management Systems

Autores: Vossen, Gottfried

Isbn-13: 9780201416046.0

Isbn-10: 201416042

Versão brasileira: Não

Titulo: Fundamentals of Database Systems

Autores: Elmasri, Ramez; Navathe, Shamkant B.

Isbn-13: 9780321122261.0

Isbn-10: 321122267

Versão brasileira: Não

Edição: 4.0

Titulo: Sistemas de Bancos de Dados

Autores: KORTH, H. F.; SILBERSCHATZ, A.; SUDARSHAN, S.

Isbn-13: 9780070310865.0

Isbn-10: 70310866

Versão brasileira: Não

Edição: 3.0

Titulo: Conceptual Database Design: an Entity-Relationship Approach

Autores: BATINI, C.; CERI, S.; NAVATHE, S.B.

Isbn-13: 9780805302448.0

Isbn-10: 805302441

Versão brasileira: Não

Edição: 1.0

Titulo: Introdução a Sistemas de Bancos de Dados

Autores: DATE, C. J.

Isbn-13: 9788535205602.0

Isbn-10: 8535205608

Versão brasileira: Sim

Edição: 7.0

Titulo: PROJETO DE BANCO DE DADOS

Autores: CARLOS ALBERTO HEUSER

Isbn-13: 9788524105906.0

Isbn-10: 8524105909

Versão brasileira: Sim

Titulo: Database: Principles, Programming, and Performance

Autores: O'NEIL, P.; O'NEIL, E.

Isbn-13: 9781558605800.0

Isbn-10: 1558605800

Versão brasileira: Não

Edição: 2.0

Titulo: SISTEMAS DE GERENCIAMENTO DE BANCO DE DADOS

Autores: JOHANNES GEHRKE, RAGHU RAMAKRISHNAN

Isbn-13: 9788577261314.0

Versão brasileira: Sim

Edição: 3.0

Titulo: Banco de Dados: Conceitos, Modelos, Gerenciadores, Projeto Lógico, Projeto Físico

Autores: SETZER,V.W.

Versão brasileira: Sim

Edição: 3.0

Titulo: Projeto e Modelagem de Banco de Dados

Autores: Teorey, Toby J.; Lightstone, Sam S.; Nadeau, Tom; Jagadish, H.V.

Isbn-13: 9780126853520.0

Isbn-10: 126853525

Versão brasileira: Não

Edição: 4.0

Titulo: First Course in Database Systems

Autores: ULLMAN, J.D., WIDOM, J.

Isbn-13: 9780138613372.0

Isbn-10: 138613370

Versão brasileira: Não

Edição: 1.0

Titulo: Data Models, Database Languages, and Database Management Systems

Autores: Vossen, Gottfried

Isbn-13: 9780201416046.0

Isbn-10: 201416042

Versão brasileira: Não

Titulo: Fundamentals of Database Systems

Autores: Elmasri, Ramez; Navathe, Shamkant B.

Isbn-13: 9780321122261.0

Isbn-10: 321122267

Versão brasileira: Não

Edição: 4.0

Titulo: Sistemas de Bancos de Dados

Autores: KORTH, H. F.; SILBERSCHATZ, A.; SUDARSHAN, S.

Isbn-13: 9780070310865.0

Isbn-10: 70310866

Versão brasileira: Não

Edição: 3.0

Titulo: Conceptual Database Design: an Entity-Relationship Approach **Autores**: BATINI, C.; CERI, S.; NAVATHE, S.B. **Isbn-13**: 9780805302448.0 **Isbn-10**: 805302441 **Versão brasileira**: Não **Edição**: 1.0
Titulo: Introdução a Sistemas de Bancos de Dados **Autores**: DATE, C. J. **Isbn-13**: 9788535205602.0 **Isbn-10**: 8535205608 **Versão brasileira**: Sim **Edição**: 7.0
Titulo: PROJETO DE BANCO DE DADOS **Autores**: CARLOS ALBERTO HEUSER **Isbn-13**: 9788524105906.0 **Isbn-10**: 8524105909 **Versão brasileira**: Sim
Titulo: Database: Principles, Programming, and Performance **Autores**: O'NEIL, P.; O'NEIL, E. **Isbn-13**: 9781558605800.0 **Isbn-10**: 1558605800 **Versão brasileira**: Não **Edição**: 2.0
Titulo: SISTEMAS DE GERENCIAMENTO DE BANCO DE DADOS **Autores**: JOHANNES GEHRKE, RAGHU RAMAKRISHNAN **Isbn-13**: 9788577261314.0 **Versão brasileira**: Sim **Edição**: 3.0
Titulo: Banco de Dados: Conceitos, Modelos, Gerenciadores, Projeto Lógico, Projeto Físico **Autores**: SETZER,V.W. **Versão brasileira**: Sim **Edição**: 3.0
Titulo: Projeto e Modelagem de Banco de Dados **Autores**: Teorey, Toby J.; Lightstone, Sam S.; Nadeau, Tom; Jagadish, H.V.

Isbn-13: 9780126853520.0

Isbn-10: 126853525

Versão brasileira: Não

Edição: 4.0

Titulo: First Course in Database Systems

Autores: ULLMAN, J.D., WIDOM, J.

Isbn-13: 9780138613372.0

Isbn-10: 138613370

Versão brasileira: Não

Edição: 1.0

Titulo: Data Models, Database Languages, and Database Management Systems

Autores: Vossen, Gottfried

Isbn-13: 9780201416046.0

Isbn-10: 201416042

Versão brasileira: Não

Titulo: Data Mining - Concepts and Techniques

Autores: HAN, J.; KAMBER, M.

Isbn-13: 9781558604896.0

Isbn-10: 1558604898

Versão brasileira: Não

Edição: 1.0

Titulo: Principles of Data Mining

Autores: HAND, D.J.; MANNILA, H.; SMYTH, P

Isbn-13: 9780262082907.0

Isbn-10: 026208290X

Versão brasileira: Não

Edição: 1.0

Titulo: Data Mining: Practical Machine Learning Tools and Techniques

Autores: Witten, Ian H.; Frank, Eibe; Hall, Mark A.; Pal, Christopher J.

Isbn-13: 9780120884070.0

Isbn-10: 120884070

Versão brasileira: Não

Titulo: Advances in Knowledge Discovery and Data Mining

Autores: FAYYAD, U.M.; PIATETSKY-SHAPIRO, G.; SMYTH, P.; UTHURUSAMY, R.

Isbn-13: 9780262560979.0

Isbn-10: 262560976

Versão brasileira: Não

Titulo: Next Generation of Data-Mining Applications

Autores: KANTARDZIC, M.; ZURADA, J.

Isbn-13: 9780471656050.0

Isbn-10: 471656054

Versão brasileira: Não

Titulo: Data Mining: Introductory and Advanced Topics

Autores: Dunham, M. H.

Isbn-13: 9780130888921.0

Isbn-10: 130888923

Versão brasileira: Não

Edição: 1.0

Titulo: Sistemas Inteligentes: Fundamentos e Aplicações

Autores: BRAGA, ANTONIO DE PADUA, REZENDE, SOLANGE OLIVEIRA, VARIOS AUTORES

Isbn-13: 9788520416839.0

Isbn-10: 8520416837

Versão brasileira: Sim

Titulo: Data Mining - Concepts and Techniques

Autores: HAN, J.; KAMBER, M.

Isbn-13: 9781558604896.0

Isbn-10: 1558604898

Versão brasileira: Não

Edição: 1.0

Titulo: Principles of Data Mining

Autores: HAND, D.J.; MANNILA, H.; SMYTH, P

Isbn-13: 9780262082907.0

Isbn-10: 026208290X

Versão brasileira: Não

Edição: 1.0

Titulo: Data Mining: Practical Machine Learning Tools and Techniques

Autores: Witten, Ian H.; Frank, Eibe; Hall, Mark A.; Pal, Christopher J.

Isbn-13: 9780120884070.0

Isbn-10: 120884070

Versão brasileira: Não

Titulo: Advances in Knowledge Discovery and Data Mining

Autores: FAYYAD, U.M.; PIATETSKY-SHAPIRO, G.; SMYTH, P.; UTHURUSAMY, R.

Isbn-13: 9780262560979.0

Isbn-10: 262560976

Versão brasileira: Não

Titulo: Next Generation of Data-Mining Applications

Autores: KANTARDZIC, M.; ZURADA, J.

Isbn-13: 9780471656050.0

Isbn-10: 471656054

Versão brasileira: Não

Titulo: Data Mining: Introductory and Advanced Topics

Autores: Dunham, M. H.

Isbn-13: 9780130888921.0

Isbn-10: 130888923

Versão brasileira: Não

Edição: 1.0

Titulo: Sistemas Inteligentes: Fundamentos e Aplicações

Autores: BRAGA, ANTONIO DE PADUA, REZENDE, SOLANGE OLIVEIRA, VARIOS AUTORES

Isbn-13: 9788520416839.0

Isbn-10: 8520416837

Versão brasileira: Sim

Titulo: The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling

Autores: KIMBALL, R., ROSS, M.

Isbn-13: 9780471200246.0

Isbn-10: 471200247

Versão brasileira: Não

Edição: 2.0

Titulo: Artigos de periódicos e congressos referentes ao processamento analítico de dados.

Titulo: Data Warehousing, Data Mining, and OLAP

Autores: Berson, Alex; Smith, Stephen J.

Isbn-13: 9780070062726.0

Isbn-10: 70062722

Versão brasileira: Não

Edição: 1.0

Titulo: Mastering Data Warehouse Design: Relational and Dimensional Techniques

Autores: Claudia Imhoff; Nicholas Galemmo; Jonathan G. Geiger

Isbn-13: 9780471324218.0

Isbn-10: 471324213

Versão brasileira: Não

Edição: 1.0

Titulo: Building the Data Warehouse

Autores: Inmon, W. H.

Isbn-13: 9780764599446.0

Isbn-10: 764599445

Versão brasileira: Não

Edição: 4.0

Titulo: The Data Warehouse ETL Toolkit: Practical Techniques for Extracting, Cleaning, Conforming, and Delivering Data

Autores: KIMBALL, R., CASERTA, J.

Isbn-13: 9780764567575.0

Isbn-10: 764567578

Versão brasileira: Não

Edição: 1.0

Titulo: The Data Warehouse Lifecycle Toolkit

Autores: KIMBALL, R., ROSS, M., THORNTHWAITE, W., MUNDY, J., BECKER, B.

Isbn-13: 9780470149775.0

Isbn-10: 470149779

Versão brasileira: Não

Edição: 2.0

Titulo: The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling

Autores: KIMBALL, R., ROSS, M.

Isbn-13: 9780471200246.0

Isbn-10: 471200247

Versão brasileira: Não

Edição: 2.0

Titulo: Artigos de periódicos e congressos referentes ao processamento analítico de dados.

Titulo: Data Warehousing, Data Mining, and OLAP

Autores: Berson, Alex; Smith, Stephen J.

Isbn-13: 9780070062726.0

Isbn-10: 70062722

Versão brasileira: Não

Edição: 1.0

Titulo: Mastering Data Warehouse Design: Relational and Dimensional Techniques

Autores: Claudia Imhoff; Nicholas Galemmo; Jonathan G. Geiger

Isbn-13: 9780471324218.0

Isbn-10: 471324213

Versão brasileira: Não

Edição: 1.0

Titulo: Building the Data Warehouse

Autores: Inmon, W. H.

Isbn-13: 9780764599446.0

Isbn-10: 764599445

Versão brasileira: Não

Edição: 4.0

Titulo: The Data Warehouse ETL Toolkit: Practical Techniques for Extracting, Cleaning, Conforming, and Delivering Data

Autores: KIMBALL, R., CASERTA, J.

Isbn-13: 9780764567575.0

Isbn-10: 764567578

Versão brasileira: Não

Edição: 1.0

Titulo: The Data Warehouse Lifecycle Toolkit

Autores: KIMBALL, R., ROSS, M., THORNTHWAITE, W., MUNDY, J., BECKER, B.

Isbn-13: 9780470149775.0

Isbn-10: 470149779

Versão brasileira: Não

Edição: 2.0

Titulo: Searching Multimedia Databases by Content

Autores: Faloutsos, Christos

Isbn-13: 9780792397779.0

Isbn-10: 792397770

Versão brasileira: Não

Titulo: Content-based object organization for efficient image retrieval in image databases

Autores: KWOK, S.H., ZHAO, J.L.

Paper: https://dl.acm.org/doi/abs/10.1016/j.dss.2006.04.013

Versão brasileira: Não

Titulo: Advanced Signature Indexing for Multimedia and Web Applications (Advances in Database Systems, 27)

Autores: Manolopoulos, Yannis; Nanopoulos, Alexandros; Tousidou, Eleni

Isbn-13: 9781461346548.0

Isbn-10: 1461346541

Versão brasileira: Não

Titulo: Content-Based Image and Video Retrieval (Multimedia Systems and Applications, 21)

Autores: Marques, Oge; Furht, Borko

Isbn-13: 9781461353386.0

Isbn-10: 1461353386

Versão brasileira: Não

Titulo: Foundations of Multidimensional and Metric Data Structures

Autores: Samet, Hanan

Isbn-13: 9780123694461.0

Isbn-10: 123694469

Versão brasileira: Não

Edição: 1.0

Titulo: Similarity Search: The Metric Space Approach (Advances in Database Systems, 32)

Autores: Zezula, Pavel - Amato, Giuseppe - Dohnal, Vlastislav - Batko, Micha

Isbn-13: 9780387291468.0

Isbn-10: 387291466

Versão brasileira: Não

Titulo: Inteligência Artificial - Uma Abordagem de Aprendizado de Máquina

Autores: KATTI FACELI PROFESSORA, ANA CAROLINA LORENA PROFESSORA, JOÃO GAMA, ANDRÉ CARLOS PONCE DE LEON FERREIRA DE CARVALHO, TIAGO AGOSTINHO DE ALMEIDA
Isbn-13: 9788521618805.0
Isbn-10: 8521618808
Versão brasileira: Sim
Edição: 1.0

Titulo: Data Science for Business: What you need to know about data mining and data-analytic thinking
Autores: Provost, F.; Fawcett, T.
Isbn-13: 9798839684010.0
Isbn-10: 1449361323
Versão brasileira: Não
Edição: 1.0

Titulo: Data Mining: Concepts and Techniques
Autores: an, Jiawei; Kamber, Micheline; Pei, Jian
Isbn-13: 9789380931913.0
Isbn-10: 9380931913
Versão brasileira: Não
Edição: 3.0

Titulo: Data Mining: Practical Machine Learning Tools and Techniques
Autores: Witten, Ian H.; Frank, Eibe; Hall, Mark A.; Pal, Christopher J.
Isbn-13: 9780123748560.0
Isbn-10: 123748569
Versão brasileira: Não

Titulo: INTRODUÇÃO AO DATAMINING: MINERAÇÃO DE DADOS
Autores: MICHAEL STEINBACH, VIPIN KUMAR
Isbn-13: 9788573937619.0
Isbn-10: 8573937610
Versão brasileira: Sim

Titulo: Cloud Computing
Autores: Rajkumar Buyya, James Broberg, Andrzej M. Goscinski.
Isbn-13: 9780470940099.0
Isbn-10: 470940093

Versão brasileira: Não

Titulo: Distributed and Cloud Computing: From Parallel Processing to the Internet of Things

Autores: Hwang, Kai; Dongarra, Jack; Fox, Geoffrey C.

Isbn-13: 9780128002049.0

Isbn-10: 128002042

Versão brasileira: Não

Edição: 1.0

Titulo: Cloud Computing and SOA Convergence in Your Enterprise: A Step-by-Step Guide

Autores: Linthicum, David S.

Isbn-13: 9780136009221.0

Isbn-10: 136009220

Versão brasileira: Não

Edição: 1.0

Titulo: Cloud Computing: Concepts, Technology & Architecture (The Pearson Service Technology Series from Thomas Erl)

Autores: Erl, Thomas; Puttini, Ricardo; Puttini, Ricardo; Mahmood, Zaigham; Mahmood, Zaigham

Isbn-13: 9780133387520.0

Isbn-10: 133387526

Versão brasileira: Não

Edição: 1.0

Titulo: Fundamentals of Database Systems

Autores: Elmasri, Ramez; Navathe, Shamkant B.

Isbn-13: 9780321122261.0

Isbn-10: 321122267

Versão brasileira: Não

Edição: 4.0

Titulo: Sistemas de Bancos de Dados

Autores: KORTH, H. F.; SILBERSCHATZ, A.; SUDARSHAN, S.

Isbn-13: 9780070310865.0

Isbn-10: 70310866

Versão brasileira: Não

Edição: 3.0

Titulo: Conceptual Database Design: an Entity-Relationship Approach

Autores: BATINI, C.; CERI, S.; NAVATHE, S.B.

Isbn-13: 9780805302448.0

Isbn-10: 805302441

Versão brasileira: Não

Edição: 1.0

Titulo: Introdução a Sistemas de Bancos de Dados

Autores: DATE, C. J.

Isbn-13: 9788535205602.0

Isbn-10: 8535205608

Versão brasileira: Sim

Edição: 7.0

Titulo: PROJETO DE BANCO DE DADOS

Autores: CARLOS ALBERTO HEUSER

Isbn-13: 9788524105906.0

Isbn-10: 8524105909

Versão brasileira: Sim

Titulo: Database: Principles, Programming, and Performance

Autores: O'NEIL, P.; O'NEIL, E.

Isbn-13: 9781558605800.0

Isbn-10: 1558605800

Versão brasileira: Não

Edição: 2.0

Titulo: SISTEMAS DE GERENCIAMENTO DE BANCO DE DADOS

Autores: JOHANNES GEHRKE, RAGHU RAMAKRISHNAN

Isbn-13: 9788577261314.0

Versão brasileira: Sim

Edição: 3.0

Titulo: Banco de Dados: Conceitos, Modelos, Gerenciadores, Projeto Lógico, Projeto Físico

Autores: SETZER,V.W.

Versão brasileira: Sim

Edição: 3.0

Titulo: Projeto e Modelagem de Banco de Dados

Autores: Teorey, Toby J.; Lightstone, Sam S.; Nadeau, Tom; Jagadish, H.V.

Isbn-13: 9780126853520.0

Isbn-10: 126853525

Versão brasileira: Não

Edição: 4.0

Titulo: First Course in Database Systems

Autores: ULLMAN, J.D., WIDOM, J.

Isbn-13: 9780138613372.0

Isbn-10: 138613370

Versão brasileira: Não

Edição: 1.0

Titulo: Data Models, Database Languages, and Database Management Systems

Autores: Vossen, Gottfried

Isbn-13: 9780201416046.0

Isbn-10: 201416042

Versão brasileira: Não

Titulo: Streaming Systems: The What, Where, When, and How of Large-Scale Data Processing

Autores: Akidau, Tyler - Chernyak, Slava - Lax, Reuven

Isbn-13: 9781491983874.0

Isbn-10: 1491983876

Versão brasileira: Não

Edição: 1.0

Titulo: Kafka - the definitive guide: Real-time data and stream processing at scale

Autores: Narkhede, Neha; Shapira, Gwen; Palino, Todd

Isbn-13: 9781491936160.0

Isbn-10: 1491936169

Versão brasileira: Não

Edição: 1.0

Titulo: Criando Microsserviços

Autores: Sam Newman

Isbn-13: 9786586057881.0

Versão brasileira: Sim

Titulo: Operadores do Kubernetes: Automatizando a plataforma de orquestração de contêineres **Autores**: Dobies, Jason; Wood, Joshua **Isbn-13**: 9786586057195.0 **Isbn-10**: 6586057191 **Versão brasileira**: Sim **Edição**: 1.0
Titulo: Notas de aula
Titulo: Designing distributed systems: patterns and paradigms for scalable, reliable services. **Autores**: Burns, Brendan **Isbn-13**: 9781491983645.0 **Isbn-10**: 1491983647 **Versão brasileira**: Não **Edição**: 1.0
Titulo: Designing data-intensive applications: The big ideas behind reliable, scalable, and maintainable system **Autores**: Kleppmann, Martin **Isbn-13**: 9781449373320.0 **Isbn-10**: 1449373321 **Versão brasileira**: Não **Edição**: 1.0
Titulo: Descomplicando o Docker **Autores**: Jeferson Fernando Noronha Vitalino, Marcus André Nunes Castro **Isbn-13**: 9788574529028.0 **Versão brasileira**: Sim **Edição**: 2.0
Titulo: Kubernetes Basico - Mergulhe no Futuro da Infraestrutura **Autores**: Brendan Burns **Isbn-13**: 9788575228241.0 **Isbn-10**: 8575228242 **Versão brasileira**: Sim
Titulo: Engenharia de Confiabilidade do Google (Português) **Isbn-13**: 9788575225172.0

Isbn-10: 8575225170

Versão brasileira: Sim

Titulo: The Kubernetes book

Autores: Poulton, Nigel

Isbn-13: 9798402153776.0

Versão brasileira: Não

Titulo: SISTEMAS DE GERENCIAMENTO DE BANCO DE DADOS

Autores: JOHANNES GEHRKE, RAGHU RAMAKRISHNAN

Isbn-13: 9788577261314.0

Versão brasileira: Sim

Edição: 3.0

Titulo: Sistemas de Bancos de Dados

Autores: KORTH, H. F.; SILBERSCHATZ, A.; SUDARSHAN, S.

Isbn-13: 9788535245356.0

Isbn-10: 8535245359

Versão brasileira: Sim

Titulo: Sistemas de Bancos de Dados

Autores: Ramez Elmasri, Shamkant B. Navathe

Isbn-13: 9788579360855.0

Isbn-10: 8579360854

Versão brasileira: Sim

Edição: 6.0

Titulo: Principles of Database and Knowledge-Base Systems

Autores: Ullman, Jeffrey D.

Isbn-13: 9780716781585.0

Isbn-10: 716781581

Versão brasileira: Não

Edição: 1.0

Titulo: Foundations of Databases: The Logical Level

Autores: Abiteboul, Serge; Hull, Richard; Vianu, Victor

Isbn-13: 9780201537710.0

Isbn-10: 201537710

Versão brasileira: Não

Edição: 1.0

Titulo: Database Systems: The Complete Book

Autores: Garcia-Molina, Hector; Ullman, Jeffrey; Widom, Jennifer

Isbn-13: 9780138155049.0

Isbn-10: 138155046

Versão brasileira: Não

Edição: 2.0

Titulo: PROJETO DE BANCO DE DADOS

Autores: CARLOS ALBERTO HEUSER

Isbn-13: 9788577804528.0

Isbn-10: 8577804526

Versão brasileira: Sim

Edição: 6.0

Titulo: Introdução a Sistemas de Bancos de Dados

Autores: DATE, C. J.

Isbn-13: 9788535212730.0

Isbn-10: 8535212736

Versão brasileira: Sim

Edição: 8.0

Titulo: SISTEMAS DE GERENCIAMENTO DE BANCO DE DADOS

Autores: JOHANNES GEHRKE, RAGHU RAMAKRISHNAN

Isbn-13: 9788577261314.0

Versão brasileira: Sim

Edição: 3.0

Titulo: PRINCIPIOS DE SISTEMAS DE BANCOS DE DADOS DISTRIBUIDOS

Autores: OZSU, M. TAWER, VALDURIEZ, PATRICK

Isbn-13: 9788535207132.0

Isbn-10: 8535207139

Versão brasileira: Sim

Edição: 1.0

Titulo: Sistemas de Bancos de Dados

Autores: Ramez Elmasri, Shamkant B. Navathe

Isbn-13: 9788579360855.0

Isbn-10: 8579360854

Versão brasileira: Sim

Edição: 6.0

Titulo: Sistemas de Bancos de Dados

Autores: KORTH, H. F.; SILBERSCHATZ, A.; SUDARSHAN, S.

Isbn-13: 9788535211078.0

Isbn-10: 8535211071

Versão brasileira: Sim

Titulo: Database Systems: The Complete Book

Autores: Garcia-Molina, Hector; Ullman, Jeffrey; Widom, Jennifer

Isbn-13: 9780138155049.0

Isbn-10: 138155046

Versão brasileira: Não

Edição: 2.0

Titulo: Architecture of a Database System (Foundations and Trends(r) in Databases)

Autores: Hellerstein, Joseph M.; Stonebraker, Michael; Hamilton, James

Isbn-13: 9781601980786.0

Isbn-10: 1601980787

Versão brasileira: Não

Titulo: Main Memory Database Systems

Autores: Faerber; Frans; Kemper; Alfons; Larson; Per-ake; Levandoski; Justin; Neumann; Thomas

Isbn-13: 9781680833249.0

Isbn-10: 1680833243

Versão brasileira: Não

Titulo: Massively Parallel Databases and Mapreduce Systems

Autores: Babu; Shivnath; Herodotou; Herodotos

Isbn-13: 9781601987501.0

Isbn-10: 1601987501

Versão brasileira: Não

Titulo: The Design and Implementation of Modern Column-Oriented Database Systems (Foundations and Trends(r) in Databases)

Autores: Abadi, Daniel; Boncz, Peter; Harizopoulos, Stavros

Isbn-13: 9781601987549.0

Isbn-10: 1601987544

Versão brasileira: Não

Titulo: Sistemas de Bancos de Dados **Autores**: Ramez Elmasri, Shamkant B. Navathe **Isbn-13**: 9788579360855.0 **Isbn-10**: 8579360854 **Versão brasileira**: Sim **Edição**: 6.0
Titulo: SISTEMAS DE GERENCIAMENTO DE BANCO DE DADOS **Autores**: JOHANNES GEHRKE, RAGHU RAMAKRISHNAN **Isbn-13**: 9788577261314.0 **Versão brasileira**: Sim **Edição**: 3.0
Titulo: Sistemas de Bancos de Dados **Autores**: KORTH, H. F.; SILBERSCHATZ, A.; SUDARSHAN, S. **Isbn-13**: 9780073523323.0 **Isbn-10**: 73523321 **Versão brasileira**: Não
Titulo: Readings in Database Systems **Autores**: Joseph M. Hellerstein and Michael Stonebraker **Isbn-13**: 9780262693141.0 **Isbn-10**: 262693143 **Versão brasileira**: Não **Edição**: 4.0
Titulo: Database Systems: The Complete Book **Autores**: Garcia-Molina, Hector; Ullman, Jeffrey; Widom, Jennifer **Isbn-13**: 9780131873254.0 **Isbn-10**: 131873253 **Versão brasileira**: Não **Edição**: 2.0
Titulo: PROJETO DE BANCO DE DADOS **Autores**: CARLOS ALBERTO HEUSER **Isbn-13**: 9788577803828.0 **Isbn-10**: 8577803821 **Versão brasileira**: Sim
Titulo: Sistemas de Bancos de Dados

Autores: Ramez Elmasri, Shamkant B. Navathe

Isbn-13: 9788579360855.0

Isbn-10: 8579360854

Versão brasileira: Sim

Edição: 6.0

Titulo: SISTEMAS DE GERENCIAMENTO DE BANCO DE DADOS

Autores: JOHANNES GEHRKE, RAGHU RAMAKRISHNAN

Isbn-13: 9788577261314.0

Versão brasileira: Sim

Edição: 3.0

Titulo: Sistemas de Bancos de Dados

Autores: KORTH, H. F.; SILBERSCHATZ, A.; SUDARSHAN, S.

Isbn-13: 9780073523323.0

Isbn-10: 73523321

Versão brasileira: Não

Titulo: Readings in Database Systems

Autores: Joseph M. Hellerstein and Michael Stonebraker

Isbn-13: 9780262693141.0

Isbn-10: 262693143

Versão brasileira: Não

Edição: 4.0

Titulo: Database Systems: The Complete Book

Autores: Garcia-Molina, Hector; Ullman, Jeffrey; Widom, Jennifer

Isbn-13: 9780131873254.0

Isbn-10: 131873253

Versão brasileira: Não

Edição: 2.0

Titulo: PROJETO DE BANCO DE DADOS

Autores: CARLOS ALBERTO HEUSER

Isbn-13: 9788577803828.0

Isbn-10: 8577803821

Versão brasileira: Sim

Titulo: Sistemas de Bancos de Dados

Autores: Ramez Elmasri, Shamkant B. Navathe

Isbn-13: 9788579360855.0

Isbn-10: 8579360854

Versão brasileira: Sim

Edição: 6.0

Titulo: Sistemas de Bancos de Dados

Autores: KORTH, H. F.; SILBERSCHATZ, A.; SUDARSHAN, S.

Isbn-13: 9788535211078.0

Isbn-10: 8535211071

Versão brasileira: Sim

Titulo: Introdução a Sistemas de Bancos de Dados

Autores: DATE, C. J.

Isbn-13: 9788535212730.0

Isbn-10: 8535212736

Versão brasileira: Sim

Edição: 8.0

Titulo: SISTEMAS DE BANCO DE DADOS: DESIGN, IMPLEMENTAÇÃO E ADMINISTRAÇÃO

Autores: ROB, Peter; CORONEL, Carlos

Isbn-13: 9788522107865.0

Isbn-10: 8522107866

Versão brasileira: Não

Titulo: Database Systems: The Complete Book

Autores: Garcia-Molina, Hector; Ullman, Jeffrey; Widom, Jennifer

Isbn-13: 9780131873254.0

Isbn-10: 131873253

Versão brasileira: Não

Edição: 2.0

Titulo: SISTEMAS DE GERENCIAMENTO DE BANCO DE DADOS

Autores: JOHANNES GEHRKE, RAGHU RAMAKRISHNAN

Isbn-13: 9788577261314.0

Versão brasileira: Sim

Edição: 3.0

Titulo: Sistemas de Bancos de Dados

Autores: Ramez Elmasri, Shamkant B. Navathe

Isbn-13: 9788579360855.0

Isbn-10: 8579360854

Versão brasileira: Sim

Edição: 6.0

Titulo: Fundamentals of Database Systems

Autores: Elmasri, Ramez; Navathe, Shamkant

Isbn-13: 9780133970777.0

Isbn-10: 133970779

Versão brasileira: Não

Edição: 7.0

Titulo: Sistemas de Bancos de Dados

Autores: KORTH, H. F.; SILBERSCHATZ, A.; SUDARSHAN, S.

Isbn-13: 9788535245356.0

Isbn-10: 8535245359

Versão brasileira: Sim

Titulo: Database Tuning: Principles, Experiments, and Troubleshooting Techniques

Autores: Shasha, Dennis; Bonnet, Philippe

Isbn-13: 9781558607538.0

Isbn-10: 1558607536

Versão brasileira: Não

Titulo: Introdução a Sistemas de Bancos de Dados

Autores: DATE, C. J.

Isbn-13: 9788535212730.0

Isbn-10: 8535212736

Versão brasileira: Sim

Titulo: Database Systems: The Complete Book

Autores: Garcia-Molina, Hector; Ullman, Jeffrey; Widom, Jennifer

Isbn-13: 9780131873254.0

Isbn-10: 131873253

Versão brasileira: Não

Edição: 2.0

Titulo: SISTEMAS DE GERENCIAMENTO DE BANCO DE DADOS

Autores: JOHANNES GEHRKE, RAGHU RAMAKRISHNAN

Isbn-13: 9788577261314.0

Versão brasileira: Sim

Edição: 3.0

Titulo: Sistemas de Bancos de Dados

Autores: Ramez Elmasri, Shamkant B. Navathe

Isbn-13: 9788579360855.0

Isbn-10: 8579360854

Versão brasileira: Sim

Edição: 6.0

Titulo: Database Tuning: Principles, Experiments, and Troubleshooting Techniques

Autores: Shasha, Dennis; Bonnet, Philippe

Isbn-13: 9781558607538.0

Isbn-10: 1558607536

Versão brasileira: Não

Titulo: Introdução a Sistemas de Bancos de Dados

Autores: DATE, C. J.

Isbn-13: 9788535212730.0

Isbn-10: 8535212736

Versão brasileira: Sim

Titulo: Sistemas de Bancos de Dados

Autores: KORTH, H. F.; SILBERSCHATZ, A.; SUDARSHAN, S.

Isbn-13: 9788535245356.0

Isbn-10: 8535245359

Versão brasileira: Sim

Titulo: Web Database Applications with PHP & MySQL, 2nd Edition

Autores: Hugh E. Williams; David Lane

Isbn-13: 9780596005436.0

Isbn-10: 596005431

Versão brasileira: Não

Edição: 2.0

Titulo: Modern Information Retrieval: The Concepts and Technology behind Search

Autores: Baeza-Yates, Ricardo; Baeza-Yates, Ricardo; Ribeiro-Neto, Berthier

Isbn-13: 9780321416919.0

Isbn-10: 321416910

Versão brasileira: Não

Edição: 2.0
Titulo: PROJETO DE BANCO DE DADOS **Autores**: CARLOS ALBERTO HEUSER **Isbn-13**: 9788577803828.0 **Isbn-10**: 8577803821 **Versão brasileira**: Sim
Titulo: Livros e artigos especializados na área de pesquisa, a serem definidos pelo Instrutor.
Titulo: IEEE Transactions on Knowledge and Data Engineering (IEEE TKDE)
Titulo: ACM Transactions on Information Systems (ACM TOIS)
Titulo: Very Large Data Base Journal (VLDB Journal)
Titulo: World Wide Web Journal
Titulo: World Wide Web Conference
Titulo: International ACM SIGIR Conference on Research and Development of Information Retrieval (ACM SIGIR)
Titulo: International Conference on Very Large Data Bases (VLDB)
Titulo: International Conference on the Management of Data (ACM SIGMOD)
Titulo: IEEE International Conference on Data Engineering (ICDE)
Titulo: IEEE Transactions on Knowledge and Data Engineering (IEEE TKDE)
Titulo: ACM Transactions on Information Systems (ACM TOIS)
Titulo: Very Large Data Base Journal (VLDB Journal)
Titulo: World Wide Web Journal
Titulo: World Wide Web Conference
Titulo: International ACM SIGIR Conference on Research and Development of Information Retrieval (ACM SIGIR)
Titulo: International Conference on Very Large Data Bases (VLDB)
Titulo: International Conference on the Management of Data (ACM SIGMOD)
Titulo: IEEE International Conference on Data Engineering (ICDE)
Titulo: Sistemas de Bancos de Dados **Autores**: Ramez Elmasri, Shamkant B. Navathe **Isbn-13**: 9788588639171.0 **Isbn-10**: 8588639173 **Versão brasileira**: Sim **Edição**: 5.0

Titulo: Fundamentals of Database Systems **Autores**: Ramez Elmasri, Shamkant B. Navathe **Isbn-13**: 9788177584769.0 **Isbn-10**: 8177584766 **Versão brasileira**: Não **Edição**: 5.0
Titulo: Sistemas de Bancos de Dados **Autores**: KORTH, H. F.; SILBERSCHATZ, A.; SUDARSHAN, S. **Isbn-13**: 9788535211078.0 **Isbn-10**: 8535211071 **Versão brasileira**: Sim
Titulo: Database Systems: The Complete Book **Autores**: Garcia-Molina, Hector; Ullman, Jeffrey; Widom, Jennifer **Isbn-13**: 9780131873254.0 **Isbn-10**: 131873253 **Versão brasileira**: Não **Edição**: 2.0
Titulo: Database Tuning: Principles, Experiments, and Troubleshooting Techniques **Autores**: Shasha, Dennis; Bonnet, Philippe **Isbn-13**: 9781558607538.0 **Isbn-10**: 1558607536 **Versão brasileira**: Não
Titulo: Introdução a Sistemas de Bancos de Dados **Autores**: DATE, C. J. **Isbn-13**: 9788535212730.0 **Isbn-10**: 8535212736 **Versão brasileira**: Sim
Titulo: Sistemas de Bancos de Dados **Autores**: KORTH, H. F.; SILBERSCHATZ, A.; SUDARSHAN, S. **Isbn-13**: 9788535211078.0 **Isbn-10**: 8535211071 **Versão brasileira**: Sim
Titulo: SISTEMAS DE GERENCIAMENTO DE BANCO DE DADOS **Autores**: JOHANNES GEHRKE, RAGHU RAMAKRISHNAN

Isbn-13: 9788577261314.0

Versão brasileira: Sim

Edição: 3.0

Titulo: Database Systems: The Complete Book

Autores: Garcia-Molina, Hector; Ullman, Jeffrey; Widom, Jennifer

Isbn-13: 9780131873254.0

Isbn-10: 131873253

Versão brasileira: Não

Edição: 2.0

Titulo: SISTEMAS DE GERENCIAMENTO DE BANCO DE DADOS

Autores: JOHANNES GEHRKE, RAGHU RAMAKRISHNAN

Isbn-13: 9788577261314.0

Versão brasileira: Sim

Edição: 3.0

Titulo: Database Tuning: Principles, Experiments, and Troubleshooting Techniques

Autores: Shasha, Dennis; Bonnet, Philippe

Isbn-13: 9781558607538.0

Isbn-10: 1558607536

Versão brasileira: Não

Titulo: Database Tuning: Principles, Experiments, and Troubleshooting Techniques

Autores: Shasha, Dennis; Bonnet, Philippe

Isbn-13: 9781558607538.0

Isbn-10: 1558607536

Versão brasileira: Não

Titulo: Postgresql - Guia do Programador

Autores: MILANI, André

Isbn-13: 9788575221570.0

Versão brasileira: Sim

Edição: 1.0

Titulo: Aprendendo Mysql

Autores: WILLIAMS, Hugh C.; M,seyed M.

Isbn-13: 9788576081470.0

Versão brasileira: Sim

Edição: 2.0

Titulo: Sistemas de Bancos de Dados **Autores**: Ramez Elmasri, Shamkant B. Navathe **Isbn-13**: 9788579360855.0 **Isbn-10**: 8579360854 **Versão brasileira**: Sim **Edição**: 6.0
Titulo: PROJETO DE BANCO DE DADOS **Autores**: CARLOS ALBERTO HEUSER **Isbn-13**: 9788577803828.0 **Isbn-10**: 8577803821 **Versão brasileira**: Sim
Titulo: Modelagem conceitual e projeto de bancos de dados **Autores**: COUGO, PAULO SERGIO **Isbn-13**: 9788535267617.0 **Versão brasileira**: Sim
Titulo: UTILIZANDO UML E PADRÕES: UMA INTRODUÇÃO À ANÁLISE E AO PROJETO ORIENTADOS A OBJETOS AO DESENVOLVIMENTO ITERATIVO **Autores**: LARMAN , CRAIG **Isbn-13**: 9788577800476.0 **Versão brasileira**: Sim
Titulo: PROJETO DE BANCO DE DADOS **Autores**: CARLOS ALBERTO HEUSER **Isbn-13**: 9788577803828.0 **Isbn-10**: 8577803821 **Versão brasileira**: Sim
Titulo: Modelagem conceitual e projeto de bancos de dados **Autores**: COUGO, PAULO SERGIO **Isbn-13**: 9788535267617.0 **Versão brasileira**: Sim
Titulo: UTILIZANDO UML E PADRÕES: UMA INTRODUÇÃO À ANÁLISE E AO PROJETO ORIENTADOS A OBJETOS AO DESENVOLVIMENTO ITERATIVO **Autores**: LARMAN , CRAIG **Isbn-13**: 9788577800476.0

Versão brasileira: Sim

Titulo: PROJETO DE BANCO DE DADOS

Autores: CARLOS ALBERTO HEUSER

Isbn-13: 9788577803828.0

Isbn-10: 8577803821

Versão brasileira: Sim

Titulo: Modelagem conceitual e projeto de bancos de dados

Autores: COUGO, PAULO SERGIO

Isbn-13: 9788535267617.0

Versão brasileira: Sim

Titulo: UTILIZANDO UML E PADRÕES: UMA INTRODUÇÃO À ANÁLISE E AO PROJETO ORIENTADOS A OBJETOS AO DESENVOLVIMENTO ITERATIVO

Autores: LARMAN , CRAIG

Isbn-13: 9788577800476.0

Versão brasileira: Sim

Titulo: Sistemas de Bancos de Dados

Autores: Ramez Elmasri, Shamkant B. Navathe

Isbn-13: 9788579360855.0

Isbn-10: 8579360854

Versão brasileira: Sim

Edição: 6.0

Titulo: Sistemas de Bancos de Dados

Autores: KORTH, H. F.; SILBERSCHATZ, A.; SUDARSHAN, S.

Isbn-13: 9788535211078.0

Isbn-10: 8535211071

Versão brasileira: Sim

Titulo: PROJETO DE BANCO DE DADOS

Autores: CARLOS ALBERTO HEUSER

Isbn-13: 9788577804528.0

Isbn-10: 8577804526

Versão brasileira: Sim

Titulo: Sistemas de Bancos de Dados

Autores: Ramez Elmasri, Shamkant B. Navathe

Isbn-13: 9788579360855.0

Isbn-10: 8579360854

Versão brasileira: Sim

Edição: 6.0

Titulo: Sistemas de Bancos de Dados

Autores: KORTH, H. F.; SILBERSCHATZ, A.; SUDARSHAN, S.

Isbn-13: 9788535211078.0

Isbn-10: 8535211071

Versão brasileira: Sim

Titulo: PROJETO DE BANCO DE DADOS

Autores: CARLOS ALBERTO HEUSER

Isbn-13: 9788577804528.0

Isbn-10: 8577804526

Versão brasileira: Sim

Titulo: Sistemas de Bancos de Dados

Autores: Ramez Elmasri, Shamkant B. Navathe

Isbn-13: 9788579360855.0

Isbn-10: 8579360854

Versão brasileira: Sim

Edição: 6.0

Titulo: Sistemas de Bancos de Dados

Autores: KORTH, H. F.; SILBERSCHATZ, A.; SUDARSHAN, S.

Isbn-13: 9788535211078.0

Isbn-10: 8535211071

Versão brasileira: Sim

Titulo: PROJETO DE BANCO DE DADOS

Autores: CARLOS ALBERTO HEUSER

Isbn-13: 9788577804528.0

Isbn-10: 8577804526

Versão brasileira: Sim

Titulo: Practical data science with R

Autores: Zumel; Mount

Isbn-13: 9789351194378.0

Isbn-10: 935119437X

Versão brasileira: Não

Titulo: An Introduction to Statistical Learning: with Applications in R **Autores**: James, Gareth; Witten, Daniela; Hastie, Trevor; Tibshirani, Robert **Isbn-13**: 9781461471370.0 **Isbn-10**: 1461471370 **Versão brasileira**: Não **Edição**: 1.0
Titulo: NDA. Notas de Aula; NDA: NDA, 2017.
Titulo: Applied Predictive Modeling **Autores**: Johnson; Kjell; Kuhn; Max **Isbn-13**: 9781461468486.0 **Isbn-10**: 1461468485 **Versão brasileira**: Não
Titulo: Foundations of Data Science **Autores**: Blum, Avrim; Hopcroft, John; Kannan, Ravindran **Isbn-13**: 9781108485067.0 **Isbn-10**: 1108485065 **Versão brasileira**: Não **Edição**: 1.0
Titulo: INTRODUÇÃO AO DATAMINING: MINERAÇÃO DE DADOS **Autores**: MICHAEL STEINBACH, VIPIN KUMAR **Isbn-13**: 9788573937619.0 **Isbn-10**: 8573937610 **Versão brasileira**: Sim
Titulo: Learning From Data A Short Course **Autores**: Abu-Mostafa, Yaser S.; Magdon-Ismail, Malik; Lin, Hsuan-Tien **Isbn-13**: 9781600490064.0 **Isbn-10**: 1600490069 **Versão brasileira**: Não
Titulo: Truthful Art, The: Data, Charts, and Maps for Communication **Autores**: Cairo, Alberto **Isbn-13**: 9780321934079.0 **Isbn-10**: 321934075 **Versão brasileira**: Não
Titulo: Sistemas de Bancos de Dados

Autores: Ramez Elmasri, Shamkant B. Navathe

Isbn-13: 9788588639171.0

Isbn-10: 8588639173

Versão brasileira: Sim

Edição: 5.0

Titulo: Introdução a Sistemas de Bancos de Dados

Autores: DATE, C. J.

Isbn-13: 9788535212730.0

Isbn-10: 8535212736

Versão brasileira: Sim

Titulo: Banco De Dados Em Aplicações Cliente-servidor

Autores: NASCIMENTO, RUBENS, NASCIMENTO, SIDLEY DIAS DA, TANAKA, ASTERIO KIYOSHI

Isbn-13: 9788573310542.0

Isbn-10: 8573310545

Versão brasileira: Sim

Titulo: GUIA DE SOBREVIVENCIA CLIENTE-SERVIDOR

Autores: ORFALI, ROBERT

Isbn-13: 9788573310290.0

Isbn-10: 8573310294

Versão brasileira: Sim

Titulo: Data Warehouse Conceitos, Tecnologias, Implementação e Gerenciamento;

Autores: SIHGH, H. S.

Isbn-13: 9788534610346.0

Isbn-10: 8534610347

Versão brasileira: Sim

Edição: 1.0

Titulo: Como Construir o Data Warehouse

Autores: INMON, W. H.

Isbn-10: 8535201416

Versão brasileira: Sim

Titulo: Sistemas de Bancos de Dados

Autores: KORTH, H. F.; SILBERSCHATZ, A.; SUDARSHAN, S.

Isbn-13: 9780070310865.0

Isbn-10: 70310866

Versão brasileira: Não

Edição: 3.0

Titulo: Sistemas de Bancos de Dados

Autores: Ramez Elmasri, Shamkant B. Navathe

Isbn-13: 9788588639171.0

Isbn-10: 8588639173

Versão brasileira: Sim

Edição: 5.0

Titulo: PROJETO DE BANCO DE DADOS

Autores: CARLOS ALBERTO HEUSER

Isbn-13: 9788577803828.0

Isbn-10: 8577803821

Versão brasileira: Sim

Titulo: Introdução a Sistemas de Bancos de Dados

Autores: DATE, C. J.

Isbn-13: 9788535212730.0

Isbn-10: 8535212736

Versão brasileira: Sim

Titulo: Database Systems: The Complete Book

Autores: Garcia-Molina, Hector; Ullman, Jeffrey; Widom, Jennifer

Isbn-13: 9780138155049.0

Isbn-10: 138155046

Versão brasileira: Não

Edição: 2.0

Titulo: First Course in Database Systems

Autores: ULLMAN, J.D., WIDOM, J.

Isbn-13: 9780130353009.0

Isbn-10: 130353000

Versão brasileira: Não

Edição: 2.0

Titulo: Implementação De Sistemas De Banco De Dados

Autores: GARCIA-MOLINA, Hector; ULLMAN, Jeffrey D.; WIDOM, Jennifer

Isbn-13: 9788535207491.0

Isbn-10: 853520749X

Versão brasileira: Sim

Edição: 1.0

Titulo: SISTEMAS DE GERENCIAMENTO DE BANCO DE DADOS

Autores: JOHANNES GEHRKE, RAGHU RAMAKRISHNAN

Isbn-13: 9788577261314.0

Versão brasileira: Sim

Edição: 3.0

Titulo: Sistemas de Bancos de Dados

Autores: KORTH, H. F.; SILBERSCHATZ, A.; SUDARSHAN, S.

Isbn-13: 9780070310865.0

Isbn-10: 70310866

Versão brasileira: Não

Edição: 3.0

Titulo: Sistemas de Bancos de Dados

Autores: Ramez Elmasri, Shamkant B. Navathe

Isbn-13: 9788588639171.0

Isbn-10: 8588639173

Versão brasileira: Sim

Edição: 5.0

Titulo: PROJETO DE BANCO DE DADOS

Autores: CARLOS ALBERTO HEUSER

Isbn-13: 9788577803828.0

Isbn-10: 8577803821

Versão brasileira: Sim

Titulo: Introdução a Sistemas de Bancos de Dados

Autores: DATE, C. J.

Isbn-13: 9788535212730.0

Isbn-10: 8535212736

Versão brasileira: Sim

Titulo: Database Systems: The Complete Book

Autores: Garcia-Molina, Hector; Ullman, Jeffrey; Widom, Jennifer

Isbn-13: 9780138155049.0

Isbn-10: 138155046

Versão brasileira: Não

Edição: 2.0

Titulo: First Course in Database Systems

Autores: ULLMAN, J.D., WIDOM, J.

Isbn-13: 9780130353009.0

Isbn-10: 130353000

Versão brasileira: Não

Edição: 2.0

Titulo: Implementação De Sistemas De Banco De Dados

Autores: GARCIA-MOLINA, Hector; ULLMAN, Jeffrey D.; WIDOM, Jennifer

Isbn-13: 9788535207491.0

Isbn-10: 853520749X

Versão brasileira: Sim

Edição: 1.0

Titulo: SISTEMAS DE GERENCIAMENTO DE BANCO DE DADOS

Autores: JOHANNES GEHRKE, RAGHU RAMAKRISHNAN

Isbn-13: 9788577261314.0

Versão brasileira: Sim

Edição: 3.0

Titulo: Inteligência Artificial

Autores: RUSSEL, S. J.; NORVING, P.

Isbn-13: 9788535211771.0

Isbn-10: 8535211772

Versão brasileira: Sim

Edição: 2.0

Titulo: Artificial Intelligence: Structures and Strategies for Complex Problem Solving

Autores: LUGER, G.F.

Isbn-13: 9780321263186.0

Isbn-10: 321263189

Versão brasileira: Não

Edição: 5.0

Titulo: The Elements of Statistical Learning: Data Mining, Inference, and Prediction

Autores: Hastie, Trevor; Tibshirani, Robert; Friedman, Jerome

Isbn-13: 9780387848570.0

Isbn-10: 387848576

Versão brasileira: Não

Edição: 2.0

Titulo: Computational intelligence: concepts to implementations

Autores: Eberhart, Russell C.; Shi, Yuhui

Isbn-13: 9781558607590.0

Isbn-10: 1558607595

Versão brasileira: Não

Edição: 1.0

Titulo: SISTEMAS INTELIGENTES: FUNDAMENTOS E APLICACOES

Autores: Solange Oliveira Rezende

Isbn-13: 9788520416839.0

Isbn-10: 8520416837

Versão brasileira: Não

Edição: 1.0

Titulo: Artificial Intelligence: a new synthesis

Autores: Nilsson, Nils J.

Isbn-13: 9781558604674.0

Isbn-10: 1558604677

Versão brasileira: Não

Edição: 1.0

Titulo: Data Mining: Practical Machine Learning Tools and Techniques

Autores: Witten, Ian H.; Frank, Eibe; Hall, Mark A.; Pal, Christopher J.

Isbn-13: 9780123748560.0

Isbn-10: 123748569

Versão brasileira: Não

Titulo: Introduction to Machine Learning

Autores: Alpaydin, Ethem

Isbn-13: 9780262012430.0

Isbn-10: 026201243X

Versão brasileira: Não

Edição: 2.0

Titulo: Inteligência Artificial

Autores: RUSSEL, S. J.; NORVING, P.

Isbn-13: 9788535211771.0

Isbn-10: 8535211772

Versão brasileira: Sim

Edição: 2.0

Titulo: Artificial Intelligence: Structures and Strategies for Complex Problem Solving

Autores: LUGER, G.F.

Isbn-13: 9780321263186.0

Isbn-10: 321263189

Versão brasileira: Não

Edição: 5.0

Titulo: The Elements of Statistical Learning: Data Mining, Inference, and Prediction

Autores: Hastie, Trevor; Tibshirani, Robert; Friedman, Jerome

Isbn-13: 9780387848570.0

Isbn-10: 387848576

Versão brasileira: Não

Edição: 2.0

Titulo: Computational intelligence: concepts to implementations

Autores: Eberhart, Russell C.; Shi, Yuhui

Isbn-13: 9781558607590.0

Isbn-10: 1558607595

Versão brasileira: Não

Edição: 1.0

Titulo: SISTEMAS INTELIGENTES: FUNDAMENTOS E APLICACOES

Autores: Solange Oliveira Rezende

Isbn-13: 9788520416839.0

Isbn-10: 8520416837

Versão brasileira: Não

Edição: 1.0

Titulo: Artificial Intelligence: a new synthesis

Autores: Nilsson, Nils J.

Isbn-13: 9781558604674.0

Isbn-10: 1558604677

Versão brasileira: Não

Edição: 1.0

Titulo: Data Mining: Practical Machine Learning Tools and Techniques

Autores: Witten, Ian H.; Frank, Eibe; Hall, Mark A.; Pal, Christopher J.

Isbn-13: 9780123748560.0

Isbn-10: 123748569

Versão brasileira: Não

Titulo: Introduction to Machine Learning

Autores: Alpaydin, Ethem

Isbn-13: 9780262012430.0

Isbn-10: 026201243X

Versão brasileira: Não

Edição: 2.0

Titulo: Sistemas de Bancos de Dados

Autores: KORTH, H. F.; SILBERSCHATZ, A.; SUDARSHAN, S.

Isbn-13: 9780070310865.0

Isbn-10: 70310866

Versão brasileira: Não

Edição: 3.0

Titulo: Sistemas de Bancos de Dados

Autores: Ramez Elmasri, Shamkant B. Navathe

Isbn-13: 9788588639171.0

Isbn-10: 8588639173

Versão brasileira: Sim

Edição: 5.0

Titulo: PROJETO DE BANCO DE DADOS

Autores: CARLOS ALBERTO HEUSER

Isbn-13: 9788577803828.0

Isbn-10: 8577803821

Versão brasileira: Sim

Titulo: Introdução a Sistemas de Bancos de Dados

Autores: DATE, C. J.

Isbn-13: 9788535212730.0

Isbn-10: 8535212736

Versão brasileira: Sim

Titulo: Database Systems: The Complete Book

Autores: Garcia-Molina, Hector; Ullman, Jeffrey; Widom, Jennifer

Isbn-13: 9780138155049.0

Isbn-10: 138155046

Versão brasileira: Não

Edição: 2.0

Titulo: First Course in Database Systems

Autores: ULLMAN, J.D., WIDOM, J.

Isbn-13: 9780130353009.0

Isbn-10: 130353000

Versão brasileira: Não

Edição: 2.0

Titulo: Implementação De Sistemas De Banco De Dados

Autores: GARCIA-MOLINA, Hector; ULLMAN, Jeffrey D.; WIDOM, Jennifer

Isbn-13: 9788535207491.0

Isbn-10: 853520749X

Versão brasileira: Sim

Edição: 1.0

Titulo: SISTEMAS DE GERENCIAMENTO DE BANCO DE DADOS

Autores: JOHANNES GEHRKE, RAGHU RAMAKRISHNAN

Isbn-13: 9788577261314.0

Versão brasileira: Sim

Edição: 3.0

Titulo: Sistemas de Bancos de Dados

Autores: KORTH, H. F.; SILBERSCHATZ, A.; SUDARSHAN, S.

Isbn-13: 9780070310865.0

Isbn-10: 70310866

Versão brasileira: Não

Edição: 3.0

Titulo: Sistemas de Bancos de Dados

Autores: Ramez Elmasri, Shamkant B. Navathe

Isbn-13: 9788588639171.0

Isbn-10: 8588639173

Versão brasileira: Sim

Edição: 5.0

Titulo: PROJETO DE BANCO DE DADOS **Autores**: CARLOS ALBERTO HEUSER **Isbn-13**: 9788577803828.0 **Isbn-10**: 8577803821 **Versão brasileira**: Sim
Titulo: Introdução a Sistemas de Bancos de Dados **Autores**: DATE, C. J. **Isbn-13**: 9788535212730.0 **Isbn-10**: 8535212736 **Versão brasileira**: Sim
Titulo: Database Systems: The Complete Book **Autores**: Garcia-Molina, Hector; Ullman, Jeffrey; Widom, Jennifer **Isbn-13**: 9780138155049.0 **Isbn-10**: 138155046 **Versão brasileira**: Não **Edição**: 2.0
Titulo: First Course in Database Systems **Autores**: ULLMAN, J.D., WIDOM, J. **Isbn-13**: 9780130353009.0 **Isbn-10**: 130353000 **Versão brasileira**: Não **Edição**: 2.0
Titulo: Implementação De Sistemas De Banco De Dados **Autores**: GARCIA-MOLINA, Hector; ULLMAN, Jeffrey D.; WIDOM, Jennifer **Isbn-13**: 9788535207491.0 **Isbn-10**: 853520749X **Versão brasileira**: Sim **Edição**: 1.0
Titulo: SISTEMAS DE GERENCIAMENTO DE BANCO DE DADOS **Autores**: JOHANNES GEHRKE, RAGHU RAMAKRISHNAN **Isbn-13**: 9788577261314.0 **Versão brasileira**: Sim **Edição**: 3.0
Titulo: The data science handbook

Autores: Cady, Field

Isbn-13: 9781119092940.0

Isbn-10: 1119092949

Versão brasileira: Não

Edição: 1.0

Titulo: Introducing data science: big data, machine learning, and more, using Python tools

Autores: Davy Cielen; Arno Meysman; Mohamed Ali

Isbn-13: 9781633430037.0

Isbn-10: 1633430030

Versão brasileira: Não

Edição: 1.0

Titulo: Introduction to Data Science: A Python Approach to Concepts, Techniques and Applications

Autores: Igual, Laura; Seguí, Santi

Isbn-13: 9783319500164.0

Isbn-10: 3319500163

Versão brasileira: Não

Edição: 1.0

Titulo: Data quality: concepts, methodology and techniques

Autores: Batini, Carlo; Scannapieco, Monica

Isbn-13: 9783540331728.0

Isbn-10: 3540331727

Versão brasileira: Não

Edição: 1.0

Titulo: Principles of data science

Autores: Ozdemir, Sinan

Isbn-13: 9781785887918.0

Isbn-10: 1785887912

Versão brasileira: Não

Edição: 1.0

Titulo: The data science design manual

Autores: Skiena, Steven S.

Isbn-13: 9783319554433.0

Isbn-10: 3319554433

Versão brasileira: Não

Edição: 1.0

Titulo: The Elements of Statistical Learning: Data Mining, Inference, and Prediction

Autores: Hastie, Trevor; Tibshirani, Robert; Friedman, Jerome

Isbn-13: 9780387848570.0

Isbn-10: 387848576

Versão brasileira: Não

Edição: 2.0

Titulo: Data Mining: Practical Machine Learning Tools and Techniques

Autores: Witten, Ian H.; Frank, Eibe; Hall, Mark A.; Pal, Christopher J.

Isbn-13: 9780123748560.0

Isbn-10: 123748569

Versão brasileira: Não

Titulo: The data science handbook

Autores: Cady, Field

Isbn-13: 9781119092940.0

Isbn-10: 1119092949

Versão brasileira: Não

Edição: 1.0

Titulo: Introducing data science: big data, machine learning, and more, using Python tools

Autores: Davy Cielen; Arno Meysman; Mohamed Ali

Isbn-13: 9781633430037.0

Isbn-10: 1633430030

Versão brasileira: Não

Edição: 1.0

Titulo: Introduction to Data Science: A Python Approach to Concepts, Techniques and Applications

Autores: Igual, Laura; Seguí, Santi

Isbn-13: 9783319500164.0

Isbn-10: 3319500163

Versão brasileira: Não

Edição: 1.0

Titulo: Data quality: concepts, methodology and techniques

Autores: Batini, Carlo; Scannapieco, Monica

Isbn-13: 9783540331728.0

Isbn-10: 3540331727

Versão brasileira: Não

Edição: 1.0

Titulo: Principles of data science

Autores: Ozdemir, Sinan

Isbn-13: 9781785887918.0

Isbn-10: 1785887912

Versão brasileira: Não

Edição: 1.0

Titulo: The data science design manual

Autores: Skiena, Steven S.

Isbn-13: 9783319554433.0

Isbn-10: 3319554433

Versão brasileira: Não

Edição: 1.0

Titulo: The Elements of Statistical Learning: Data Mining, Inference, and Prediction

Autores: Hastie, Trevor; Tibshirani, Robert; Friedman, Jerome

Isbn-13: 9780387848570.0

Isbn-10: 387848576

Versão brasileira: Não

Edição: 2.0

Titulo: Data Mining: Practical Machine Learning Tools and Techniques

Autores: Witten, Ian H.; Frank, Eibe; Hall, Mark A.; Pal, Christopher J.

Isbn-13: 9780123748560.0

Isbn-10: 123748569

Versão brasileira: Não

Titulo: Sistemas de Bancos de Dados

Autores: Ramez Elmasri, Shamkant B. Navathe

Isbn-13: 9788579360855.0

Isbn-10: 8579360854

Versão brasileira: Sim

Edição: 6.0

Titulo: Sistemas de Bancos de Dados

Autores: KORTH, H. F.; SILBERSCHATZ, A.; SUDARSHAN, S.

Isbn-13: 9788535245356.0

Isbn-10: 8535245359

Versão brasileira: Sim

Edição: 6.0

Titulo: Database Systems: Design, Implementation, and Management

Autores: Coronel, Carlos; Morris, Steven; Rob, Peter

Isbn-13: 9781111969608.0

Isbn-10: 1111969604

Versão brasileira: Não

Edição: 8.0

Titulo: Database Reliability Engineering: Designing and Operating Resilient Database Systems

Autores: Campbell, Laine; Majors, Charity

Isbn-13: 9781491925942.0

Isbn-10: 1491925949

Versão brasileira: Não

Edição: 1.0

Titulo: Introdução a Sistemas de Bancos de Dados

Autores: DATE, C. J.

Isbn-13: 9788535212730.0

Isbn-10: 8535212736

Versão brasileira: Sim

Titulo: Modelagem e Banco de Dados

Autores: CABRAL, Alex Lima; SANTANA FILHO, Ozeas Vieira

Isbn-13: 9788539622221.0

Isbn-10: 853962222X

Versão brasileira: Sim

Edição: 2.0

Titulo: Projeto e Modelagem de Banco de Dados

Autores: Teorey, Toby J.; Lightstone, Sam S.; Nadeau, Tom; Jagadish, H.V.

Isbn-13: 9780126853520.0

Isbn-10: 126853525

Versão brasileira: Não

Edição: 4.0

Titulo: Banco de Dados

Autores: ALVES, William Pereira

Versão brasileira: Sim

Edição: 1.0

Titulo: Inteligência artificial

Autores: RUSSEL, S.; NORVIG, P.

Isbn-13: 9788535251418.0

Versão brasileira: Sim

Edição: 3.0

Titulo: Inteligência Artificial: Ferramentas e Teorias

Autores: BITTENCOURT, Guilherme

Isbn-13: 9788532801388.0

Isbn-10: 8532801382

Versão brasileira: Sim

Edição: 1.0

Titulo: Neural Networks and Learning Machines

Autores: Haykin, Simon

Isbn-13: 9780131471399.0

Isbn-10: 131471392

Versão brasileira: Não

Edição: 3.0

Titulo: O que é inteligência artificial

Autores: João de Fernandes Teixeira

Isbn-10: 8577290131

Versão brasileira: Sim

Titulo: Inteligencia Artificial

Autores: George F. Luger

Isbn-13: 9788581435503.0

Isbn-10: 8581435505

Versão brasileira: Sim

Edição: 6.0

Titulo: Artificial Intelligence

Autores: WINSTON, P. H.

Isbn-13: 9780201533774.0

Isbn-10: 201533774

Versão brasileira: Não

Edição: 3.0

Titulo: Artificial Intelligence

Autores: RICH, Elaine; KNIGHT, Kevin

Isbn-13: 9780071008945.0

Isbn-10: 71008942

Versão brasileira: Não

Edição: 2.0

Titulo: Lógica Para Ciência Da Computação

Autores: SOUZA, João Nunes de

Isbn-13: 9788535229615.0

Isbn-10: 8535229612

Versão brasileira: Sim

Edição: 2.0

Titulo: Database Reliability Engineering: Designing and Operating Resilient Database Systems

Autores: Campbell, Laine; Majors, Charity

Isbn-13: 9781491925942.0

Isbn-10: 1491925949

Versão brasileira: Não

Edição: 1.0

Titulo: Sistemas de Bancos de Dados

Autores: Ramez Elmasri, Shamkant B. Navathe

Isbn-13: 9788579360855.0

Isbn-10: 8579360854

Versão brasileira: Sim

Edição: 6.0

Titulo: Sistemas de Bancos de Dados

Autores: KORTH, H. F.; SILBERSCHATZ, A.; SUDARSHAN, S.

Isbn-13: 9788535245356.0

Isbn-10: 8535245359

Versão brasileira: Sim

Titulo: Database Systems: Design, Implementation, and Management

Autores: Coronel, Carlos; Morris, Steven; Rob, Peter

Isbn-13: 9781111969608.0

Isbn-10: 1111969604

Versão brasileira: Não

Edição: 8.0

Titulo: Introdução a Sistemas de Bancos de Dados

Autores: DATE, C. J.

Isbn-13: 9788535212730.0

Isbn-10: 8535212736

Versão brasileira: Sim

Edição: 8.0

Titulo: Principles of Big Data: Preparing, Sharing, and Analyzing Complex Information

Autores: Berman, Jules J.

Isbn-13: 9780124045767.0

Isbn-10: 124045766

Versão brasileira: Não

Edição: 1.0

Titulo: Data Warehousing in the Age of Big Data

Autores: Krishnan, Krish

Isbn-13: 9780124058910.0

Isbn-10: 124058914

Versão brasileira: Não

Edição: 1.0

Titulo: SISTEMAS DE BANCO DE DADOS: DESIGN, IMPLEMENTAÇÃO E ADMINISTRAÇÃO

Autores: ROB, Peter; CORONEL, Carlos

Isbn-13: 9788522107865.0

Isbn-10: 8522107866

Versão brasileira: Não

Titulo: Inteligência artificial

Autores: RUSSEL, S.; NORVIG, P.

Isbn-13: 9788535251418.0

Versão brasileira: Sim

Edição: 3.0

Titulo: Inteligência Artificial: Ferramentas e Teorias

Autores: BITTENCOURT, Guilherme

Isbn-13: 9788532801388.0

Isbn-10: 8532801382

Versão brasileira: Sim

Edição: 1.0

Titulo: Neural Networks and Learning Machines

Autores: Haykin, Simon

Isbn-13: 9780131471399.0

Isbn-10: 131471392

Versão brasileira: Não

Edição: 3.0

Titulo: Pattern Recognition and Machine Learning

Autores: Bishop, Christopher M.

Isbn-13: 9780387310732.0

Isbn-10: 387310738

Versão brasileira: Não

Titulo: Pattern Classification

Autores: Duda, Richard O.; Hart, Peter E.; Stork, David G.

Isbn-13: 9780471056690.0

Isbn-10: 471056693

Versão brasileira: Não

Edição: 2.0

Titulo: The Elements of Statistical Learning: Data Mining, Inference, and Prediction

Autores: Hastie, Trevor; Tibshirani, Robert; Friedman, Jerome

Isbn-13: 9780387848570.0

Isbn-10: 387848576

Versão brasileira: Não

Edição: 2.0

Titulo: Machine Learning

Autores: MITCHELL, T. M.

Isbn-13: 9789384761165.0

Isbn-10: 9384761168

Versão brasileira: Não

Titulo: Pattern Recognition and Machine Learning

Autores: GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A.

Titulo: Data mining: conceitos, técnicas, algoritmos, orientações e aplicações

Autores: Ronaldo Goldschmidt; Emmanuel Passos; Eduardo Bezerra

Isbn-13: 9788535278231.0

Isbn-10: 8535278230

Versão brasileira: Sim

Titulo: Data Mining - Concepts and Techniques

Autores: HAN, J.; KAMBER, M.

Isbn-13: 9781558604896.0

Isbn-10: 1558604898

Versão brasileira: Não

Edição: 1.0

Titulo: Data Preprocessing in Data Mining

Autores: García; Salvador; Luengo; Julián; Herrera; Francisco

Isbn-13: 9783319377315.0

Isbn-10: 3319377310

Versão brasileira: Não

Titulo: Inteligência Competitiva Em Tempos De Big Data

Autores: Elisabeth Gomes; Fabiane Braga

Isbn-13: 9788550806389.0

Isbn-10: 8550806382

Versão brasileira: Sim

Titulo: Principles of Big Data: Preparing, Sharing, and Analyzing Complex Information

Autores: Berman, Jules J.

Isbn-13: 9780124045767.0

Isbn-10: 124045766

Versão brasileira: Não

Edição: 1.0

Titulo: Data Warehousing in the Age of Big Data

Autores: Krishnan, Krish

Isbn-13: 9780124058910.0

Isbn-10: 124058914

Versão brasileira: Não **Edição**: 1.0
Titulo: Sistemas de Bancos de Dados **Autores**: Ramez Elmasri, Shamkant B. Navathe **Isbn-13**: 9788579360855.0 **Isbn-10**: 8579360854 **Versão brasileira**: Sim **Edição**: 6.0
Titulo: Sistemas de Bancos de Dados **Autores**: KORTH, H. F.; SILBERSCHATZ, A.; SUDARSHAN, S. **Isbn-13**: 9788535245356.0 **Isbn-10**: 8535245359 **Versão brasileira**: Sim
Titulo: Data mining: conceitos, técnicas, algoritmos, orientações e aplicações **Autores**: Ronaldo Goldschmidt; Emmanuel Passos; Eduardo Bezerra **Isbn-13**: 9788535278231.0 **Isbn-10**: 8535278230 **Versão brasileira**: Sim
Titulo: Data Mining - Concepts and Techniques **Autores**: HAN, J.; KAMBER, M. **Isbn-13**: 9781558604896.0 **Isbn-10**: 1558604898 **Versão brasileira**: Não **Edição**: 1.0
Titulo: Data Preprocessing in Data Mining **Autores**: García; Salvador; Luengo; Julián; Herrera; Francisco **Isbn-13**: 9783319377315.0 **Isbn-10**: 3319377310 **Versão brasileira**: Não
Titulo: Inteligência Competitiva Em Tempos De Big Data **Autores**: Elisabeth Gomes; Fabiane Braga **Isbn-13**: 9788550806389.0 **Isbn-10**: 8550806382 **Versão brasileira**: Sim

Titulo: Principles of Big Data: Preparing, Sharing, and Analyzing Complex Information

Autores: Berman, Jules J.

Isbn-13: 9780124045767.0

Isbn-10: 124045766

Versão brasileira: Não

Edição: 1.0

Titulo: Data Warehousing in the Age of Big Data

Autores: Krishnan, Krish

Isbn-13: 9780124058910.0

Isbn-10: 124058914

Versão brasileira: Não

Edição: 1.0

Titulo: Sistemas de Bancos de Dados

Autores: Ramez Elmasri, Shamkant B. Navathe

Isbn-13: 9788579360855.0

Isbn-10: 8579360854

Versão brasileira: Sim

Edição: 6.0

Titulo: Sistemas de Bancos de Dados

Autores: KORTH, H. F.; SILBERSCHATZ, A.; SUDARSHAN, S.

Isbn-13: 9788535245356.0

Isbn-10: 8535245359

Versão brasileira: Sim

Titulo: Sistemas de Bancos de Dados

Autores: Ramez Elmasri, Shamkant B. Navathe

Isbn-13: 9788579360855.0

Isbn-10: 8579360854

Versão brasileira: Sim

Edição: 6.0

Titulo: PROJETO DE BANCO DE DADOS

Autores: CARLOS ALBERTO HEUSER

Isbn-13: 9788577804528.0

Isbn-10: 8577804526

Versão brasileira: Sim
Titulo: SQL: The Complete Reference **Autores**: James R. Groff; Paul N. Weinberg **Isbn-13**: 9780072225594.0 **Isbn-10**: 72225599 **Versão brasileira**: Não **Edição**: 2.0
Titulo: Sistemas de Bancos de Dados **Autores**: KORTH, H. F.; SILBERSCHATZ, A.; SUDARSHAN, S. **Isbn-13**: 9788535211078.0 **Isbn-10**: 8535211071 **Versão brasileira**: Sim
Titulo: First Course in Database Systems **Autores**: ULLMAN, J.D., WIDOM, J. **Isbn-13**: 9780136006374.0 **Isbn-10**: 013600637X **Versão brasileira**: Não **Edição**: 3.0

D

Bibliografia deduplicada

Livros
Titulo: sistemas de bancos de dados **Autores**: ramez elmasri, shamkant b. navathe **Isbn-13**: 9788588639171.0 **Isbn-10**: 8588639173.0
Titulo: sistemas de bancos de dados **Autores**: ramez elmasri, shamkant b. navathe **Isbn-13**: 9788579360855.0 **Isbn-10**: 8579360854.0
Titulo: projeto de banco de dados **Autores**: carlos alberto heuser **Isbn-13**: 9788577803828.0 **Isbn-10**: 8577803821.0
Titulo: introdução a sistemas de bancos de dados **Autores**: date, c. j. **Isbn-13**: 9788535205602.0 **Isbn-10**: 8535205608.0
Titulo: introdução a sistemas de bancos de dados **Autores**: date, c. j. **Isbn-13**: 9788535212730.0

Isbn-10: 8535212736.0

Titulo: bancos de dados em aplicações cliente/servidor

Autores: nascimento, rubens, nascimento, sidley dias da, tanaka, asterio kiyoshi

Isbn-10: 8573310545.0

Titulo: database administration: the complete guide to practices and procedures

Autores: mullins, craig s.

Isbn-13: 9780201741292.0

Isbn-10: 201741296.0

Titulo: database tuning: principles, experiments, and troubleshooting techniques

Autores: shasha, dennis; bonnet, philippe

Isbn-13: 9781558607538.0

Isbn-10: 1558607536.0

Titulo: estatística básica

Autores: morettin,pedro alberto,bussab,wilton de oliveira

Isbn-13: 9788502207998.0

Isbn-10: 8502207997.0

Titulo: sistemas de bancos de dados

Autores: ramez elmasri, shamkant b. navathe

Isbn-13: 9786550110512.0

Titulo: sistemas de bancos de dados

Autores: korth, h. f.; silberschatz, a.; sudarshan, s.

Isbn-13: 9788535245356.0

Isbn-10: 8535245359.0

Titulo: projeto de banco de dados

Autores: carlos alberto heuser

Isbn-13: 9788577804528.0

Isbn-10: 8577804526.0

Titulo: sistemas de gerenciamento de banco de dados
Autores: johannes gehrke, raghu ramakrishnan
Isbn-13: 9788577261314.0

Titulo: sistemas de banco de dados: design, implementação e administração
Autores: carlos coronel
Isbn-13: 9788522107865.0
Isbn-10: 8522107866.0

Titulo: data mining: practical machine learning tools and techniques
Autores: witten, ian h.; frank, eibe; hall, mark a.; pal, christopher j.
Isbn-13: 9780128042915.0
Isbn-10: 128042915.0

Titulo: an introduction to categorical data analysis
Autores: agresti, alan
Isbn-13: 9780471226185.0
Isbn-10: 471226181.0

Titulo: mining of massive datasets
Autores: leskovec, jure
Isbn-13: 9781107077232.0
Isbn-10: 1107077230.0

Titulo: data mining and data warehousing: principles and practical techniques
Autores: bhatia, parteek
Isbn-13: 9781108727747.0
Isbn-10: 1108727743.0

Titulo: data science and big data analytics: discovering, analyzing, visualizing and presenting data
Autores: emc
Isbn-13: 9781118876138.0

Titulo: hands-on machine learning with scikit-learn and tensorflow: concepts, tools, and techniques to build intelligent systems **Autores**: géron, aurélien **Isbn-13**: 9781491962299.0 **Isbn-10**: 1491962291.0
Titulo: introduction to data science: data analysis and prediction algorithms with r **Autores**: irizarry, r.a **Isbn-13**: 9780367357986.0 **Isbn-10**: 367357984.0
Titulo: applied predictive modeling **Autores**: johnson; kjell; kuhn; max **Isbn-13**: 9781461468486.0 **Isbn-10**: 1461468485.0
Titulo: fundamentals of database systems **Autores**: elmasri, ramez; navathe, shamkant b. **Isbn-13**: 9780321122261.0 **Isbn-10**: 321122267.0
Titulo: sistemas de bancos de dados **Autores**: korth, h. f.; silberschatz, a.; sudarshan, s. **Isbn-13**: 9780070310865.0 **Isbn-10**: 70310866.0
Titulo: conceptual database design: an entity-relationship approach **Autores**: batini, c.; ceri, s.; navathe, s.b. **Isbn-13**: 9780805302448.0 **Isbn-10**: 805302441.0
Titulo: projeto de banco de dados **Autores**: carlos alberto heuser

Isbn-13: 9788524105906.0

Isbn-10: 8524105909.0

Titulo: database: principles, programming, and performance

Autores: o'neil, p.; o'neil, e.

Isbn-13: 9781558605800.0

Isbn-10: 1558605800.0

Titulo: projeto e modelagem de banco de dados

Autores: teorey, toby j.; lightstone, sam s.; nadeau, tom; jagadish, h.v.

Isbn-13: 9780126853520.0

Isbn-10: 126853525.0

Titulo: first course in database systems

Autores: ullman, j.d., widom, j.

Isbn-13: 9780138613372.0

Isbn-10: 138613370.0

Titulo: data models, database languages, and database management systems

Autores: vossen, gottfried

Isbn-13: 9780201416046.0

Isbn-10: 201416042.0

Titulo: data mining - concepts and techniques

Autores: han, j.; kamber, m.

Isbn-13: 9781558604896.0

Isbn-10: 1558604898.0

Titulo: principles of data mining

Autores: hand, d.j.; mannila, h.; smyth, p

Isbn-13: 9780262082907.0

Titulo: data mining: practical machine learning tools and techniques

Autores: witten, ian h.; frank, eibe; hall, mark a.; pal, christopher j.

Isbn-13: 9780120884070.0

Isbn-10: 120884070.0

Titulo: advances in knowledge discovery and data mining

Autores: fayyad, u.m.; piatetsky-shapiro, g.; smyth, p.; uthurusamy, r.

Isbn-13: 9780262560979.0

Isbn-10: 262560976.0

Titulo: next generation of data-mining applications

Autores: kantardzic, m.; zurada, j.

Isbn-13: 9780471656050.0

Isbn-10: 471656054.0

Titulo: data mining: introductory and advanced topics

Autores: dunham, m. h.

Isbn-13: 9780130888921.0

Isbn-10: 130888923.0

Titulo: sistemas inteligentes: fundamentos e aplicações

Autores: braga, antonio de padua, rezende, solange oliveira, varios autores

Isbn-13: 9788520416839.0

Isbn-10: 8520416837.0

Titulo: the data warehouse toolkit: the complete guide to dimensional modeling

Autores: kimball, r., ross, m.

Isbn-13: 9780471200246.0

Isbn-10: 471200247.0

Titulo: data warehousing, data mining, and olap

Autores: berson, alex; smith, stephen j.

Isbn-13: 9780070062726.0

Isbn-10: 70062722.0

Titulo: mastering data warehouse design: relational and dimensional techniques

Autores: claudia imhoff; nicholas galemmo; jonathan g. geiger

Isbn-13: 9780471324218.0

Isbn-10: 471324213.0

Titulo: building the data warehouse

Autores: inmon, w. h.

Isbn-13: 9780764599446.0

Isbn-10: 764599445.0

Titulo: the data warehouse etl toolkit: practical techniques for extracting, cleaning, conforming, and delivering data

Autores: kimball, r., caserta, j.

Isbn-13: 9780764567575.0

Isbn-10: 764567578.0

Titulo: the data warehouse lifecycle toolkit

Autores: kimball, r., ross, m., thornthwaite, w., mundy, j., becker, b.

Isbn-13: 9780470149775.0

Isbn-10: 470149779.0

Titulo: searching multimedia databases by content

Autores: faloutsos, christos

Isbn-13: 9780792397779.0

Isbn-10: 792397770.0

Titulo: advanced signature indexing for multimedia and web applications (advances in database systems, 27)

Autores: manolopoulos, yannis; nanopoulos, alexandros; tousidou, eleni

Isbn-13: 9781461346548.0

Isbn-10: 1461346541.0

Titulo: content-based image and video retrieval (multimedia systems and applications, 21)

Autores: marques, oge; furht, borko

Isbn-13: 9781461353386.0

Isbn-10: 1461353386.0

Titulo: foundations of multidimensional and metric data structures

Autores: samet, hanan

Isbn-13: 9780123694461.0

Isbn-10: 123694469.0

Titulo: similarity search: the metric space approach (advances in database systems, 32)

Autores: zezula, pavel - amato, giuseppe - dohnal, vlastislav - batko, micha

Isbn-13: 9780387291468.0

Isbn-10: 387291466.0

Titulo: inteligência artificial - uma abordagem de aprendizado de máquina

Autores: katti faceli professora, ana carolina lorena professora, joão gama, andré carlos ponce de leon ferreira de carvalho, tiago agostinho de almeida

Isbn-13: 9788521618805.0

Isbn-10: 8521618808.0

Titulo: data science for business: what you need to know about data mining and data-analytic thinking

Autores: provost, f.; fawcett, t.

Isbn-13: 9798839684010.0

Isbn-10: 1449361323.0

Titulo: data mining: concepts and techniques

Autores: an, jiawei; kamber, micheline; pei, jian

Isbn-13: 9789380931913.0

Isbn-10: 9380931913.0

Titulo: data mining: practical machine learning tools and techniques

Autores: witten, ian h.; frank, eibe; hall, mark a.; pal, christopher j.

Isbn-13: 9780123748560.0

Isbn-10: 123748569.0

Titulo: introdução ao datamining: mineração de dados

Autores: michael steinbach, vipin kumar

Isbn-13: 9788573937619.0

Isbn-10: 8573937610.0

Titulo: cloud computing **Autores**: rajkumar buyya, james broberg, andrzej m. goscinski. **Isbn-13**: 9780470940099.0 **Isbn-10**: 470940093.0
Titulo: distributed and cloud computing: from parallel processing to the internet of things **Autores**: hwang, kai; dongarra, jack; fox, geoffrey c. **Isbn-13**: 9780128002049.0 **Isbn-10**: 128002042.0
Titulo: cloud computing and soa convergence in your enterprise: a step-by-step guide **Autores**: linthicum, david s. **Isbn-13**: 9780136009221.0 **Isbn-10**: 136009220.0
Titulo: cloud computing: concepts, technology & architecture (the pearson service technology series from thomas erl) **Autores**: erl, thomas; puttini, ricardo; puttini, ricardo; mahmood, zaigham; mahmood, zaigham **Isbn-13**: 9780133387520.0 **Isbn-10**: 133387526.0
Titulo: streaming systems: the what, where, when, and how of large-scale data processing **Autores**: akidau, tyler - chernyak, slava - lax, reuven **Isbn-13**: 9781491983874.0 **Isbn-10**: 1491983876.0
Titulo: kafka - the definitive guide: real-time data and stream processing at scale **Autores**: narkhede, neha; shapira, gwen; palino, todd **Isbn-13**: 9781491936160.0 **Isbn-10**: 1491936169.0
Titulo: criando microsserviços **Autores**: sam newman

Isbn-13: 9786586057881.0

Titulo: operadores do kubernetes: automatizando a plataforma de orquestração de contêineres
Autores: dobies, jason; wood, joshua
Isbn-13: 9786586057195.0
Isbn-10: 6586057191.0

Titulo: designing distributed systems: patterns and paradigms for scalable, reliable services.
Autores: burns, brendan
Isbn-13: 9781491983645.0
Isbn-10: 1491983647.0

Titulo: designing data-intensive applications: the big ideas behind reliable, scalable, and maintainable system
Autores: kleppmann, martin
Isbn-13: 9781449373320.0
Isbn-10: 1449373321.0

Titulo: descomplicando o docker
Autores: jeferson fernando noronha vitalino, marcus andré nunes castro
Isbn-13: 9788574529028.0

Titulo: kubernetes basico - mergulhe no futuro da infraestrutura
Autores: brendan burns
Isbn-13: 9788575228241.0
Isbn-10: 8575228242.0

Titulo: engenharia de confiabilidade do google (português)
Isbn-13: 9788575225172.0
Isbn-10: 8575225170.0

Titulo: the kubernetes book

Autores: poulton, nigel

Isbn-13: 9798402153776.0

Titulo: principles of database and knowledge-base systems

Autores: ullman, jeffrey d.

Isbn-13: 9780716781585.0

Isbn-10: 716781581.0

Titulo: foundations of databases: the logical level

Autores: abiteboul, serge; hull, richard; vianu, victor

Isbn-13: 9780201537710.0

Isbn-10: 201537710.0

Titulo: database systems: the complete book

Autores: garcia-molina, hector; ullman, jeffrey; widom, jennifer

Isbn-13: 9780138155049.0

Isbn-10: 138155046.0

Titulo: principios de sistemas de bancos de dados distribuidos

Autores: ozsu, m. tawer, valduriez, patrick

Isbn-13: 9788535207132.0

Isbn-10: 8535207139.0

Titulo: sistemas de bancos de dados

Autores: korth, h. f.; silberschatz, a.; sudarshan, s.

Isbn-13: 9788535211078.0

Isbn-10: 8535211071.0

Titulo: architecture of a database system (foundations and trends(r) in databases)

Autores: hellerstein, joseph m.; stonebraker, michael; hamilton, james

Isbn-13: 9781601980786.0

Isbn-10: 1601980787.0

Titulo: main memory database systems

Autores: faerber; frans; kemper; alfons; larson; per-ake; levandoski; justin; neumann; thomas
Isbn-13: 9781680833249.0
Isbn-10: 1680833243.0

Titulo: massively parallel databases and mapreduce systems
Autores: babu; shivnath; herodotou; herodotos
Isbn-13: 9781601987501.0
Isbn-10: 1601987501.0

Titulo: the design and implementation of modern column-oriented database systems (foundations and trends(r) in databases)
Autores: abadi, daniel; boncz, peter; harizopoulos, stavros
Isbn-13: 9781601987549.0
Isbn-10: 1601987544.0

Titulo: sistemas de bancos de dados
Autores: korth, h. f.; silberschatz, a.; sudarshan, s.
Isbn-13: 9780073523323.0
Isbn-10: 73523321.0

Titulo: readings in database systems
Autores: joseph m. hellerstein and michael stonebraker
Isbn-13: 9780262693141.0
Isbn-10: 262693143.0

Titulo: database systems: the complete book
Autores: garcia-molina, hector; ullman, jeffrey; widom, jennifer
Isbn-13: 9780131873254.0
Isbn-10: 131873253.0

Titulo: fundamentals of database systems
Autores: elmasri, ramez; navathe, shamkant
Isbn-13: 9780133970777.0

Isbn-10: 133970779.0

Titulo: web database applications with php & mysql, 2nd edition
Autores: hugh e. williams; david lane
Isbn-13: 9780596005436.0
Isbn-10: 596005431.0

Titulo: modern information retrieval: the concepts and technology behind search
Autores: baeza-yates, ricardo; baeza-yates, ricardo; ribeiro-neto, berthier
Isbn-13: 9780321416919.0
Isbn-10: 321416910.0

Titulo: fundamentals of database systems
Autores: ramez elmasri, shamkant b. navathe
Isbn-13: 9788177584769.0
Isbn-10: 8177584766.0

Titulo: postgresql - guia do programador
Autores: milani, andré
Isbn-13: 9788575221570.0

Titulo: aprendendo mysql
Autores: williams, hugh c.; m,seyed m.
Isbn-13: 9788576081470.0

Titulo: modelagem conceitual e projeto de bancos de dados
Autores: cougo, paulo sergio
Isbn-13: 9788535267617.0

Titulo: utilizando uml e padrões: uma introdução à análise e ao projeto orientados a objetos ao desenvolvimento iterativo
Autores: larman , craig
Isbn-13: 9788577800476.0

Titulo: practical data science with r

Autores: zumel; mount

Isbn-13: 9789351194378.0

Titulo: an introduction to statistical learning: with applications in r

Autores: james, gareth; witten, daniela; hastie, trevor; tibshirani, robert

Isbn-13: 9781461471370.0

Isbn-10: 1461471370.0

Titulo: foundations of data science

Autores: blum, avrim; hopcroft, john; kannan, ravindran

Isbn-13: 9781108485067.0

Isbn-10: 1108485065.0

Titulo: learning from data a short course

Autores: abu-mostafa, yaser s.; magdon-ismail, malik; lin, hsuan-tien

Isbn-13: 9781600490064.0

Isbn-10: 1600490069.0

Titulo: truthful art, the: data, charts, and maps for communication

Autores: cairo, alberto

Isbn-13: 9780321934079.0

Isbn-10: 321934075.0

Titulo: banco de dados em aplicações cliente-servidor

Autores: nascimento, rubens, nascimento, sidley dias da, tanaka, asterio kiyoshi

Isbn-13: 9788573310542.0

Isbn-10: 8573310545.0

Titulo: guia de sobrevivencia cliente-servidor

Autores: orfali, robert

Isbn-13: 9788573310290.0

Isbn-10: 8573310294.0

Titulo: data warehouse conceitos, tecnologias, implementação e gerenciamento;

Autores: sihgh, h. s.

Isbn-13: 9788534610346.0

Isbn-10: 8534610347.0

Titulo: first course in database systems

Autores: ullman, j.d., widom, j.

Isbn-13: 9780130353009.0

Isbn-10: 130353000.0

Titulo: implementação de sistemas de banco de dados

Autores: garcia-molina, hector; ullman, jeffrey d.; widom, jennifer

Isbn-13: 9788535207491.0

Titulo: inteligência artificial

Autores: russel, s. j.; norving, p.

Isbn-13: 9788535211771.0

Isbn-10: 8535211772.0

Titulo: artificial intelligence: structures and strategies for complex problem solving

Autores: luger, g.f.

Isbn-13: 9780321263186.0

Isbn-10: 321263189.0

Titulo: the elements of statistical learning: data mining, inference, and prediction

Autores: hastie, trevor; tibshirani, robert; friedman, jerome

Isbn-13: 9780387848570.0

Isbn-10: 387848576.0

Titulo: computational intelligence: concepts to implementations

Autores: eberhart, russell c.; shi, yuhui

Isbn-13: 9781558607590.0

Isbn-10: 1558607595.0

Titulo: artificial intelligence: a new synthesis

Autores: nilsson, nils j.

Isbn-13: 9781558604674.0

Isbn-10: 1558604677.0

Titulo: introduction to machine learning **Autores**: alpaydin, ethem **Isbn-13**: 9780262012430.0
Titulo: the data science handbook **Autores**: cady, field **Isbn-13**: 9781119092940.0 **Isbn-10**: 1119092949.0
Titulo: introducing data science: big data, machine learning, and more, using python tools **Autores**: davy cielen; arno meysman; mohamed ali **Isbn-13**: 9781633430037.0 **Isbn-10**: 1633430030.0
Titulo: introduction to data science: a python approach to concepts, techniques and applications **Autores**: igual, laura; seguí, santi **Isbn-13**: 9783319500164.0 **Isbn-10**: 3319500163.0
Titulo: data quality: concepts, methodology and techniques **Autores**: batini, carlo; scannapieco, monica **Isbn-13**: 9783540331728.0 **Isbn-10**: 3540331727.0
Titulo: principles of data science **Autores**: ozdemir, sinan **Isbn-13**: 9781785887918.0 **Isbn-10**: 1785887912.0
Titulo: the data science design manual **Autores**: skiena, steven s. **Isbn-13**: 9783319554433.0 **Isbn-10**: 3319554433.0

Titulo: database systems: design, implementation, and management

Autores: coronel, carlos; morris, steven; rob, peter

Isbn-13: 9781111969608.0

Isbn-10: 1111969604.0

Titulo: database reliability engineering: designing and operating resilient database systems

Autores: campbell, laine; majors, charity

Isbn-13: 9781491925942.0

Isbn-10: 1491925949.0

Titulo: modelagem e banco de dados

Autores: cabral, alex lima; santana filho, ozeas vieira

Isbn-13: 9788539622221.0

Titulo: inteligência artificial

Autores: russel, s.; norvig, p.

Isbn-13: 9788535251418.0

Titulo: inteligência artificial: ferramentas e teorias

Autores: bittencourt, guilherme

Isbn-13: 9788532801388.0

Isbn-10: 8532801382.0

Titulo: neural networks and learning machines

Autores: haykin, simon

Isbn-13: 9780131471399.0

Isbn-10: 131471392.0

Titulo: inteligencia artificial

Autores: george f. luger

Isbn-13: 9788581435503.0

Isbn-10: 8581435505.0

Titulo: artificial intelligence

Autores: winston, p. h.

Isbn-13: 9780201533774.0

Isbn-10: 201533774.0

Titulo: artificial intelligence

Autores: rich, elaine; knight, kevin

Isbn-13: 9780071008945.0

Isbn-10: 71008942.0

Titulo: lógica para ciência da computação

Autores: souza, joão nunes de

Isbn-13: 9788535229615.0

Isbn-10: 8535229612.0

Titulo: principles of big data: preparing, sharing, and analyzing complex information

Autores: berman, jules j.

Isbn-13: 9780124045767.0

Isbn-10: 124045766.0

Titulo: data warehousing in the age of big data

Autores: krishnan, krish

Isbn-13: 9780124058910.0

Isbn-10: 124058914.0

Titulo: pattern recognition and machine learning

Autores: bishop, christopher m.

Isbn-13: 9780387310732.0

Isbn-10: 387310738.0

Titulo: pattern classification

Autores: duda, richard o.; hart, peter e.; stork, david g.

Isbn-13: 9780471056690.0

Isbn-10: 471056693.0

Titulo: machine learning

Autores: mitchell, t. m.

Isbn-13: 9789384761165.0

Isbn-10: 9384761168.0

Titulo: data mining: conceitos, técnicas, algoritmos, orientações e aplicações

Autores: ronaldo goldschmidt; emmanuel passos; eduardo bezerra

Isbn-13: 9788535278231.0

Isbn-10: 8535278230.0

Titulo: data preprocessing in data mining

Autores: garcía; salvador; luengo; julián; herrera; francisco

Isbn-13: 9783319377315.0

Isbn-10: 3319377310.0

Titulo: inteligência competitiva em tempos de big data

Autores: elisabeth gomes; fabiane braga

Isbn-13: 9788550806389.0

Isbn-10: 8550806382.0

Titulo: SQL :the complete reference.

Autores: Groff, James R.; Weinberg, Paul N

Isbn-13: 9780072225594.0

Isbn-10: 72225599.0

E

CAPÍTULOS SELECIONADOS DA BIBLIOGRAFIA LEVANTADA

Livros
Banco de Dados em aplicações Cliente-Servidor • introdução 1 • A Tecnologia Utilizada no Ambiente Cliente-Servidor • A Arquitetura da Aplicação • O Modelo Relacional e a Linguagem SQL • Gerência de Transações • Projeto de Bancos de Dados Relacionais • Bancos de Dados Distribuídos • Acesso a Dados Distribuídos por Sistemas de Banco de Dados Independentes • Tópicos Correntes em Banco de Dados
PROJETO DE BANCO DE DADOS 2008 Heuser

- introdução: banco de dados e modelos de banco de dados
- abordagem entidade-relacionamento
- construindo modelos ER
- abordagem relacional
- transformações entre modelos
- engenharia reversa de arquivos e normalização

SQL :the complete reference

- introduction
- A quick tour of SQL
- SQL in Perspective
- SQL Basics
- Simple Queries
- Multitable Queries
- Summary Queries
- Subqueries and Query Expressions
- Database Updates
- Data Integrity
- Transaction Processing
- Creating a Database
- Views
- SQL Security
- The System Catalog
- Embedded SQL
- Dynamic SQL
- SQL APIs
- Database Processing and Stored Procedures
- SQL and Data Warehousing
- SQL and Application Servers
- SQL Networking and Distributed Databases
- SQL and Objects
- SQL and XML
- The future of SQL

Sistemas de banco de dados 6ª ed

- bancos de dados e usuarios de banco de dados
- conceitos e arquitetura do sistema de banco de dados
- O modelo de dados relacional e as restrições em bancos de dados relacionais
- SQL básica
- Mais SQL: Consultas complexas, triggers, views e modificação de esquema
- Álgebra e cálculo relacional
- Modelagem de dados usando o modelo Entidade-Relacionamento (ER).
- O modelo Entidade-Relacionamento Estendido (EER)
- Projeto de banco de dados relacional por mapeamento ER e EER para relacional
- Metodologia prática de projeto de banco de dados e uso de diagramas UML
- Bancos de dados de objeto e objeto-relacional
- XML: Extensible Markup Language
- Introdução às técnicas de programação SQL
- Programação de banco de dados Web usando PHP
- Fundamentos de dependências funcionais e normalização para bancos de dados relacionais
- Algoritmos de projeto de banco de dados relacional e demais dependências
- Armazenamento de disco, estruturas de arquivo básicas e hashing
- Estruturas de indexação para arquivos
- Algoritmos para processamento e otimização de consulta
- Projeto físico e ajuste de banco de dados
- Introdução aos conceitos e teoria de processamento de transações
- Técnicas de controle de concorrência
- Técnicas de recuperação de banco de dados
- Segurança de banco de dados

Sistemas de banco de dados 7ª ed

- estrategias para processamento de consulta
- otimização de consulta
- bancos de dados nosql e sistemas de armazenamento big data
- tecnologias big data baseadas em mapreduce e hadoop

advanced signature indexing for multimedia and web applications (advances in database systems, 27)

- introduction
- <https://link.springer.com/chapter/10.1007/978-1-4419-8636-8_2>
- <https://link.springer.com/chapter/10.1007/978-1-4419-8636-8_3>
- <https://link.springer.com/chapter/10.1007/978-1-4419-8636-8_4>
- <https://link.springer.com/chapter/10.1007/978-1-4419-8636-8_5>
- <https://link.springer.com/chapter/10.1007/978-1-4419-8636-8_6>
- <https://link.springer.com/chapter/10.1007/978-1-4419-8636-8_7>
- <https://link.springer.com/chapter/10.1007/978-1-4419-8636-8_8>
- <https://link.springer.com/chapter/10.1007/978-1-4419-8636-8_9>

an introduction to categorical data analysis

- introduction
- analyzing contingence tables
- generalized linear models
- logistic regression
- building and applying logistic regression models
- multicategory logit models
- loglinear models for contingency tables and counts
- models for matched pairs
- marginal modeling of correlated, clustered responses
- random effects: generalized linear mixed models
- classification and smoothing
- a historical tour of categorical data analysis

an introduction to statistical learning: with applications in r

- Introduction
- Statistical Learning
- Linear Regression
- Classification
- Resampling Methods
- Linear Model Selection and Regularization
- Moving Beyond Linearity
- Tree-Based Methods
- Support Vector Machines
- Unsupervised Learning

applied predictive modeling

- introduction
- a short tour of the predictive modeling process
- data pre-processing
- over-fitting and model tuning
- measuring performance in regression models
- linear regression and its cousins
- nonlinear regression models
- regression trees and rule-based models
- a summary of solutibility models
- case study: compressive strength of concrete mixtures
- measuring performance in classification models
- discriminant analysis and other linear classification models
- nonlinear classification models
- classification trees and rule-based models
- a summary of grant application models
- remedies for severe class imbalance
- case study: job scheduling
- measuring predictor importance
- an introduction of features selection
- factors that can affect model performance

aprendendo mysql

- Introdução
- Instalação do MySQL
- Uso do MySQL Monitor
- Modelagem e Projeto de Banco de Dados
- SQL Básico
- Trabalhando com Tabelas do Banco de Dados
- Consultas Avançadas
- Obtendo mais do MySQL: Importação de dados e outros tópicos
- Gerenciamento de Usuários e Permissões de Acesso
- Backup e Restore
- Uso de um Arquivo de Opções
- Configuração e Ajuste do Servidor
- Aplicações Web
- PHP
- Perl
- Usando Perl com MySQL
- Fornecendo páginas Perl para a Web

architecture of a database system (foundations and trends(r) in databases)

- architecture of a database system (foundations and trends(r) in databases) (paper)

artificial intelligence

- what is artificial inteligence
- problems, problem spaces, and search
- Heuristic Search techniques
- knowledge representation issues
- using predicate logic
- representing knowledge using rules
- symbolic reasoning under uncertainty
- statistical reasoning
- weak slot-and-filter structures
- strong slog-and-filter structures
- knowledge representation summary
- game playing
- planning
- understanding
- Natural Language Processing
- parallel and distributed AI
- learning
- connectionist models
- common sense
- expert systems
- perception and action
- fuzzy logic systems
- genetic algorithms: copying nature's approaches
- artificial immune systems
- prolog-the natural language of artificial intelligence

artificial intelligence: a new synthesis

- introduction
- stimulus-response agents
- neural networks
- machine evolution
- state machines
- robot vision
- agents that plan
- uninformed search
- Heuristic Search
- planning, acting and learning
- alternative search forumulations and applications
- adversarial search
- the propositional calculus
- resolution in the propositional calculus
- The Predicate Calculus
- resolution in the predicate calculus
- knowledge-base systems
- representing commonsense knowledge
- Reasoning with Uncertain information
- learning and acting with bayes nets
- the situation calculus
- planning
- multiple agents
- communication among agents
- agent architechtures

artificial intelligence: structures and strategies for complex problem solving

- AI: Early History and Applications
- The Predicate Calculus
- Structures and Strategies for State Space Search
- Heuristic Search
- Stochastic Methods
- Building Control Algorithms for State Space Search
- Knowledge Representation
- Strong Method Problem Solving
- Reasoning in Uncertain Situations
- Machine Learning: Symbol-Based
- Machine Learning: Connectionist
- Machine Learning: Social and Emergent
- Automated Reasoning
- Understanding Natural Language
- An Introduction to Prolog
- An Introduction to Lisp
- Artificial Intelligence as Empirical Enquiry

building the data warehouse

- Evolution of Decision Support Systems
- The Data Warehouse Environment
- The Data Warehouse and Design
- Granularity in the Data Warehouse
- The Data Warehouse and Technology
- The Distributed Data Warehouse
- Executive Information Systems and the Data Warehouse
- External Data and the Data Warehouse
- Migration to the Architected Environment
- The Data Warehouse and the Web
- Unstructured Data and the Data Warehouse
- The Really Large Data Warehouse
- The Relational and the Multidimensional Models as a Basis for Database Design
- Data Warehouse Advanced Topics
- Cost-Justification and Return on Investment for a Data Warehouse
- The Data Warehouse and the ODS
- Corporate Information Compliance and Data Warehousing
- The End-User Community
- Data Warehouse Design Review Checklist

cloud computing

- Introduction to Cloud Computing
- Migrating into a cloud
- Enriching the 'integration as a service' paradigm for the cloud era
- The enterprise cloud computing paradigm
- Virtual Machines Provisioning and migration services
- on the management of virtual machines for cloud infrastructures
- Enhancing cloud computing environments using a cluster as a service
- secure distributed data
- Aneka - integration of private and public clouds
- Cometcloud: an autonomic cloud engine
- T-Systems' Cloud-based solutions for business applications
- Workflow engine for clouds
- Understanding scientific applications for cloud environments
- The mapreduce programming mode and implementations
- An architecture for federated cloud computing
- SLA management in cloud
- Performance prediction for hpc on clouds
- Best pratices in architecting cloud applications in the AWS cloud
- Massively multiplayer online game hosting on cloud resources
- Building Content delivery networks using clouds
- Resource cloud mashups
- Organizational readiness and change management in the cloud age
- Data security in the cloud
- Legal issues in cloud computing
- Achieving production readiness for cloud services

cloud computing and soa convergence in your enterprise: a step-by-step guide

- Where we are, how we got here, and how to fix it
- Reaching for the Clouds
- Defining the clouds for the enterprise
- Making the business case for clouds
- Working from your data to the clouds
- Working from your service to the clouds
- Bringing governance to the clouds
- Testing from SOA to the Clouds
- Defining candidate data, services, and processes for the clouds
- Making the move to cloud computing
- Moving onward

cloud computing: concepts, technology & architecture (the pearson service technology series from thomas erl)

- Introduction
- Case Study background
- understanding cloud computing
- fundamental concepts and models
- cloud-enabling technology
- fundamental cloud security
- Cloud infrastructure mechanisms
- Specialized cloud mechanisms
- Cloud management mechanisms
- Cloud security mechanisms
- Fundamental cloud architechtures
- Advanced cloud architechtures
- Specialized cloud architechtures
- Cloud delivery model considerations
- Cost metrics and pricing models
- Service quality metrics and SLAs

computational intelligence: concepts to implementations

- foundations
- computational inteligence
- evolutionary computation concepts and paradigms
- evolutionary computation implementations
- neural Network concepts and paradigms
- neural network implementations
- fuzzy systems concepts and paradigms
- fuzzy systems implementations
- computational inteligence implementations
- performance metrics
- analysis explanation

conceptual database design: an entity-relationship approach

- An Introduction to Database Design
- Data Modeling Concepts
- methodologies for conceptual design
- view design
- view integration
- improving the quality of database schema
- schema documentation and maintenance
- functional analysis using the dataflow model
- joint data and functional analysis
- case study
- High-Level Logical Design Using the Entity-Relationship Model
- Logical Design for the Relational Model
- Logical Design for the Network Model
- Logical Design for the Hierarchical Model
- Database Design Tools

content-based image and video retrieval (multimedia systems and applications, 21)

- <https://link.springer.com/chapter/10.1007/978-1-4615-0987-5_1>
- <https://link.springer.com/chapter/10.1007/978-1-4615-0987-5_2>
- <https://link.springer.com/chapter/10.1007/978-1-4615-0987-5_3>
- <https://link.springer.com/chapter/10.1007/978-1-4615-0987-5_4>
- <https://link.springer.com/chapter/10.1007/978-1-4615-0987-5_5>
- <https://link.springer.com/chapter/10.1007/978-1-4615-0987-5_6>

criando microsserviços

- o que são microsserviços
- como modelar microsserviços
- dividindo o sistema monolitico
- estilos de comunicação entre microsserviços
- implementando a comunicação entre microsserviços
- fluxo de trabalho
- construção
- implantação
- testes
- da monitoração a observabilidade
- segurança
- Resiliencia
- escalabilidade
- interfaces de usuario
- estruturas organizacionais
- o arquiteto em evolução

data mining - concepts and techniques

- Introduction
- getting to know your data
- data preprocessing
- Data Warehousing and Online Analytical Processing
- Data Cube Technology
- Mining Frequent Patterns, Associations, and Correlations: Basic Concepts and Methods
- Advanced Pattern Mining
- Classification: Basic Concepts
- Classification: Advanced Methods
- Cluster Analysis: Basic Concepts and Methods
- Advanced Cluster Analysis
- Outlier Detection
- Data Mining Trends and Research Frontiers

data mining and data warehousing: principles and practical techniques

- Beginning with Machine Learning
- Introduction to Data Mining
- Beginning with Weka and R Language
- Data Preprocessing
- Classification
- Implementing Classification in Weka and R
- Cluster Analysis
- Implementing Clustering with Weka and R
- Association Mining
- Implementing Association Mining with Weka and R
- Web Mining and Search Engines
- Data Warehouse
- Data Warehouse Schema
- Online Analytical Processing
- Big Data and NoSQL

data mining: conceitos, técnicas, algoritmos, orientações e aplicações

- Introdução
- Processo de KDD: Conceitos Básicos
- Etapas do Processo de KDD
- Tarefas de KDD
- Métodos de Mineração de Dados
- Metodologia de KDD
- Mineração de Grafos
- Big Data
- Mineração de Dados Educacionais
- Exemplos de Aplicação de KDD
- Considerações finais

data mining: introductory and advanced topics

- Introduction
- Related Concepts
- Data Mining Techniques
- Classification
- Clustering
- Association Rules
- Web Mining
- Spatial Mining
- Temporal Mining

data mining: practical machine learning tools and techniques

- what's it all about
- input: concepts, instances, and attributtes
- output: knowledge representation
- algorithms: the basic methods
- credibility: evaluating whats been learned
- implementations: real machine learning schemes
- data transformations
- ensemble learning
- moving on: applications and beyond
- introduction to weka
- the explorer
- the knowledge flow interface
- the experimenter
- the command-line interface
- embedded machine learning
- writing new learning schemes
- tutorial exercises for the weka explorer

data preprocessing in data mining

- Introduction
- Data Sets and Proper Statistical Analysis of Data Mining Techniques
- Data Preparation Basic Models
- Dealing with Missing Values
- Dealing with Noisy Data
- Data Reduction
- Feature Selection
- Instance Selection
- Discretization
- A Data Mining Software Package Including Data Preparation and Reduction: KEEL

data quality: concepts, methodology and techniques

- Introduction to Data Quality
- Data Quality Dimensions
- Models for Data Quality
- Activities and Techniques for Data Quality: Generalities
- Object Identification
- Data Quality Issues in Data Integration Systems
- Methodologies for Data Quality Measurement and Improvement
- Tools for Data Quality
- Open Problems

data science and big data analytics: discovering, analyzing, visualizing and presenting data

- introduction to big data analytics
- data analytics lifecycle
- review of basic data analytic methods using R
- advanced analytical theory and methods: clustering
- advanced analytical theory and methods: association rules
- advanced analytical theory and methods: regression
- advanced analytical theory and methods: classification
- advanced analytical theory and methods: time series analysis
- advanced analytical theory and methods: text analysis
- advanced analytics - technology and tools: mapreduce and hadoop
- advanced analytics - technology and tools: database analytics
- the endgame, or putting it all together

data science for business: what you need to know about data mining and data-analytic thinking

- Introduction: Data-Analytic Thinking
- Business problems and data science solutions
- Introduction to Predictive Modeling: From correlation to Supervised Segmentation
- Fitting a model to data
- Overfitting and its avoidance
- Similarity, neighbors and Clusters
- Decision Analytic Thinking 1: What is a good model?
- Visualizing model performance
- Evidence and Probabilities
- Representing and Mining Text
- Decision Analytic Thinking 2: Toward Analytical Engineering
- Other data science tasks and techniques
- Data science and Business Strategy
- Conclusion

data warehousing in the age of big data

- Introduction to Big Data
- Working with Big Data
- Big Data Processing Architectures
- Introducing Big Data Technologies
- Big Data Driving Business Value
- Data Warehousing Revisited
- Reengineering the Data Warehouse
- Workload Management in the Data Warehouse
- New Technologies Applied to Data Warehousing
- Integration of Big Data and Data Warehousing
- Data-Driven Architecture for Big Data
- Information Management and Life Cycle for Big Data
- Big Data Analytics, Visualization, and Data Scientists
- Implementing the Big Data – Data Warehouse – Real-Life Situations

database administration: the complete guide to practices and procedures

- What Is a DBA?
- <https://learning.oreilly.com/library/view/database-administration-the/0201741296/ch02.html>
- <https://learning.oreilly.com/library/view/database-administration-the/0201741296/ch03.html>
- <https://learning.oreilly.com/library/view/database-administration-the/0201741296/ch04.html>
- <https://learning.oreilly.com/library/view/database-administration-the/0201741296/ch05.html>
- <https://learning.oreilly.com/library/view/database-administration-the/0201741296/ch06.html>
- <https://learning.oreilly.com/library/view/database-administration-the/0201741296/ch07.html?_gl=1*1d23ajm*_ga*MTMyMjM0NjA3LjE2Nzc2MjgxNjY.*_ga_092EL089CH*MTY3NzYyODE2NS4xLjEuMTY3NzYyODM5Mi42MC4wLjA.>
- <https://learning.oreilly.com/library/view/database-administration-the/0201741296/ch08.html>
- <https://learning.oreilly.com/library/view/database-administration-the/0201741296/ch09.html>
- <https://learning.oreilly.com/library/view/database-administration-the/0201741296/ch10.html?_gl=1*607pk8*_ga*MTMyMjM0NjA3LjE2Nzc2MjgxNjY.*_ga_092EL089CH*MTY3NzYyODE2NS4xLjEuMTY3NzYyODQyNi4yNi4wLjA.>
- <https://learning.oreilly.com/library/view/database-administration-the/0201741296/ch11.html>
- <https://learning.oreilly.com/library/view/database-administration-the/0201741296/ch12.html>
- <https://learning.oreilly.com/library/view/database-administration-the/0201741296/ch13.html>
- <https://learning.oreilly.com/library/view/database-administration-the/0201741296/ch14.html>

database reliability engineering: designing and operating resilient database systems

- Introducing Database Reliability Engineering
- Service-Level Management
- Risk Management
- Operational Visibility
- Infrastructure Engineering
- Infrastructure Management
- Backup and Recovery
- Release Management
- Security
- Data Storage, Indexing, and Replication
- Datastore Field Guide
- A Data Architecture Sampler
- Making the Case For DBRE

database systems: the complete book

- the worlds of database systems
- the relational model of data
- design theory for relational databases
- high-level database model
- algebraic and logical query languages
- the database language SQL
- constraints and triggers
- views and indexes
- SQL in server environment
- advanced topics in relational databases
- the semistructured-data model
- programming languages for XML
- secondary storage management
- index structures
- query execution
- the query compiler
- coping with system failures
- concurrency control
- more about transaction management
- parallel and distributed databases
- information integration
- data mining
- database systems and the internet

database tuning: principles, experiments, and troubleshooting techniques

- Basic Principles
- tuning the guts
- index tuning
- tuning relational systems
- communicating with outside
- case studies from wall streets
- troubleshooting
- tuning e-commerce applications
- celko on data warehouses: techniques, successes, and mistakes
- data warehouse tuning
- real-time databases
- transaction chopping
- time series, especially for finance
- understanding access plans
- configuration parameters

descomplicando o docker

- o que é container
- o que é docker
- instalando docker
- executando e administrando containers docker
- configurando CPU e memoria para os meus containers
- meu primeiro e tosco Dockerfile
- entendendo volumes
- criando e gerenciando imagens
- compartilhando as imagens
- gerenciando a rede dos containers
- controlando o daemon do docker
- Docker Machine
- Docker Swarm
- Docker Secrets
- Docker Compose

designing data-intensive applications: the big ideas behind reliable, scalable, and maintainable system

- Reliable, scalable and maintainable applications
- Data Model and Query Languages
- Storage and Retrieval
- Encoding and Evolution
- Replication
- Partitioning
- Transactions
- The Trouble with Distributed Systems
- Consistency and Consensus
- Batch processing
- Stream processing
- The future of Data systems

designing distributed systems: patterns and paradigms for scalable, reliable services.

- introduction
- the sidecar pattern
- ambassadors
- adapters
- replicated load-balanced services
- shared services
- scatter/gather
- functions and event-driven processing
- ownership election
- work queue systems
- event driven batch processing
- coordinated batch processing
- Conclusion: a new beginning

distributed and cloud computing: from parallel processing to the internet of things

- Distributed system models and enabling technologies
- Computer clusters for scalable parallel computing
- Virtual Machines and virtualization of clusters and data centers
- cloud platform architecture over virtualized data centers
- service-oriented architectures for distributed computing
- cloud programming and software environments
- Grid computing systems and resource management
- Peer-to-peer computing and overlay networks
- Ubiquitous clouds and the internet of things

engenharia de confiabilidade do google (português)

- introdução
- o ambiente de produção do Google do ponto de vista de um SRE
- aceitando riscos
- objetivos do nivel de serviço
- eliminando tarefas penosas
- monitorando sistemas distribuidos
- a evolução da automação no Google
- engenharia de release
- simplicidade
- alertas praticos a partir de dados de series temporais
- de plantão
- resolvendo problemas de modo eficiente
- respostas e emergencias
- administrando incidentes
- cultura de postmortem: aprendendo com o fracasso
- monitorando interrupções de serviços
- testes voltados a confiabilidade
- engenharia de software em SRE
- distribuição de carga no frontend
- distribuição de carga no datacenter
- tratando sobrecarga
- tratando falhas em cascata
- administrando estados criticos: consenso distribuido para confiabilidade
- escalonamento periodico e distribuido com cron
- pipelines de processamento de dados

estatistica basica

- preliminares
- resumo dos dados
- medidas-resumo
- analise bidimensional
- probabilidades
- variaveis aleatorias discretas
- variaveis aleatorias continuas
- variaveis aleatorias multidimensionais
- noções de simulação
- introdução a inferencia estatistica
- estimação
- testes de hipotese
- inferencia para duas populações
- analise de aderencia e assosiação
- inferencia para varias populações
- regressão linear simples

first course in database systems

- the worlds of database systems
- the relational model of data
- design theory for relational databases
- high-level database models
- algebraic and logical query languages
- the database language SQL
- constraints and triggers
- views and indexes
- sql in a server environment
- advanced topics in relational databases
- the semistructured-data model

foundations of data science

- introduction
- high-dimensional space
- best-fit subspaces and singular value decomposition
- random walks and markov chains
- machine learning
- algorithms for massive data problems: streaming, sketching, and sampling
- clustering
- random graphs
- topic models, nonnegative matrix factorization, hidden markov models, and graphical models
- other topics
- wavelets
- background material

foundations of databases: the logical level

- Database Systems
- Theoretical Background
- The Relational Model
- Conjunctive Queries
- Adding Negation: Algebra and Calculus
- Static Analysis and Optimization
- Notes on Practical Languages
- Functional and Join Dependency
- Inclusion Dependency
- A Larger Perspective
- Design and Dependencies
- Datalog
- Evaluation of Datalog
- Recursion and Negation
- Negation in Datalog
- Sizing Up Languages
- First Order, Fixpoint and While
- Highly Expressive Languages
- Incomplete Information
- Complex Values
- Object Databases
- Dynamic Aspects

foundations of multidimensional and metric data structures

- Multidimensional point data
- Object-based and image-based representations
- Intervals and Small rectangles
- High-dimensional data

fundamentals of database systems

- databases and database users
- database system concepts and architecture
- the relational data model and relational database constraints
- basic sql
- more SQL: complex queries, triggers, views, and schema modification
- the relational algebra and relational calculus
- data modeling using the Entity-Relationship (ER) model
- the enhanced entity-relationship (EER) model
- relational database by ER and EER-to-relational Mapping
- pratical database design methodology and use of UML diagrams
- object and object-relational databases
- XML: Extensible Markup Language
- introduction to SQL programming technique
- web database programming using PHP
- basics of functional dependencies and normalization for relational databases
- relational database design algorithms and further dependencies
- Disk storage, basic file structures, and hashing
- indexing structures for file
- algorithms for query processing
- physical database design and tuning
- introduction to transaction processing concepts and theory
- concurrency control techniques
- database recovery techniques
- database security
- distributed database

hands-on machine learning with scikit-learn and tensorflow: concepts, tools, and techniques to build intelligent systems

- o cenario do aprendizado de maquina
- projeto de aprendizado de maquina de ponta a ponta
- classificação
- treinando modelos
- maquians de vetores de suporte
- arvores de decisão
- ensemble learning e florestas aleatorias
- redução de dimensionalidade
- em pleno funcionamento com o TensorFlow
- introdução a redes neurais artificiais
- treinando redes neurais profundas
- distribuindo o tensorflow por dispositivos e servidores
- redes neurais convolucionais (CNN)
- redes neurais recorrentes (RNN)
- autoencoders
- aprendizado por reforço

inteligencia artificial

- IA: historia e aplicaçoes
- calculo de predicados
- estruturas e estrategias para busca em espaço de estados
- busca heuristica
- metodos estocasticos
- construção de algoritmos de controle para busca em espaço de estados
- representação do conhecimento
- solução de problemas por metodo forte
- raciocinando em situações incertas
- aprendizado de maquina: simbolico
- aprendizado de maquina: conexionista
- aprendizado de maquina: genetico e emergente
- aprendizado de maquina: probabilistico
- raciocinio automatizado
- compreensão da linguagem natural
- inteligencia artificial como investigação empirica

inteligência artificial

- introdução
- agentes inteligentes
- resolução de problemas por meio de busca
- além da busca classica
- busca competitiva
- problemas de satisfação de restrições
- agentes logics
- logica de primeira ordem
- inferencia em logica de primeira ordem
- planejamento classico
- planejamento e ação no mundo real
- representação de conhecimento
- quantificando a incerteza
- raciocinio probabilistico
- raciocinio probabilistico temporal
- tomada de decisões simples
- tomada de decisões complexas
- aprendendo a partir de exemplos
- conhecimento em aprendizagem
- aprendizagem em modelos probabilisticos
- aprendizagem por reforço
- processamento de linguagem natural
- linguagem natural para comunicação
- percepção
- robotica

inteligência artificial - uma abordagem de aprendizado de máquina

- Introdução
- Analise de dados
- pré-processamento de dados
- metodos baseados em distancias
- metodos probabilisticos
- metodos simbolicos
- metodos conexionistas
- metodos de maximização de margens
- modelos multiplos preditivos
- avaliação de modelos preditivos
- mineração de padrões frequentes
- Analise de agrupamentos
- Algoritmo de agrupamento
- modelos multiplos descritivos
- avaliação de modelos descritivos
- Aprendizador em fluxos continuos de dados
- Aprendizado de maquina automatizado
- Decomposição de problemas multiclasse
- classificação multirrotulo
- classificação hierarquica
- computação natural
- analise de redes sociais
- categorização de textos
- Agronegocios
- Analise de Sentimento

inteligência competitiva em tempos de big data

- introdução
- a inteligencia competitiva
- inteligencia competitiva e o processo de tomada de decisão
- inteligencia competitiva no contexto de big data: como usar e porque usar
- praticando
- conclusao

introducing data science: big data, machine learning, and more, using python tools

- data science in the big data world
- the data science process
- machine learning
- handling large data on a single computer
- first steps in big data
- join the nosql movement
- the rise of graph databases
- text mining and text analytics
- data visualization to the end user

introduction to data science: a python approach to concepts, techniques and applications

- Introduction to Data Science
- Toolboxes for Data Scientists
- Descriptive Statistics
- Statistical Inference
- Supervised Learning
- Regression Analysis
- Unsupervised Learning
- Network Analysis
- Recommender Systems
- Statistical Natural Language Processing for Sentiment Analysis
- Parallel Computing

introduction to data science: data analysis and prediction algorithms with r

- getting started with R and RStudio
- R basics
- programming basics
- the tydeverse
- importing data
- data.table
- introduction to data visualization
- ggplot2
- visualizing data distributions
- data visualization in pratice
- data visualization principles
- introduction to statistics with R
- summary statistics
- probability
- random variables
- statistical inference
- statistical models
- regression
- linear models
- association is not causation
- introduction to data wrangling
- reshaping data
- joining tables
- web scraping
- string processing

introduction to machine learning

- Introduction
- Supervised Learning
- Bayesian Decision Theory
- Parametric Methods
- Multivariate Methods
- Dimensionality Reduction
- Clustering
- Nonparametric Methods
- Decision Trees
- Linear Discrimination
- Multilayer Perceptrons
- Local Models
- Kernel Machines
- Graphical Models
- Hidden Markov Models
- Bayesian Estimation
- Combining Multiple Learners
- Reinforcement Learning
- Design and Analysis of Machine Learning Experiments

introdução a sistemas de bancos de dados

- visão geral do gerenciamento de dados
- arquitetura de sistemas de banco de dados
- introdução aos bancos de dados relacionais
- uma introdução a SQL
- tipos
- relações
- algebra relacional
- calculo relacional
- integridade
- visões
- dependencias funcionais
- normalização avançada I: 1FN, 2FN, 3FN, FNBC
- normalização avançada II: formas normais maiores
- Modelagem semântica
- recuperação
- concorrencia
- segurança
- otimização
- falta de informações
- herança de tipos
- bancos de dados distribuidos
- Apoio a decisão
- sistemas baseados em logica
- sistemas baseados em logica
- bancos de dados relacional/objeto

kafka - the definitive guide: real-time data and stream processing at scale

- Meet Kafka
- instaling kafka
- Kafka producers: writing messages to kafka
- Kafka consumers: reading data from kafka
- Kafka internals
- Reliable data delivery
- Building data pipelines
- Cross-cluster data mirroring
- Administering Kafka
- Monitoring Kafka
- Stream processing

kubernetes basico - mergulhe no futuro da infraestrutura

- Introdução
- criando e executando containers
- implantação de um cluster kubernetes
- comandos comuns do kubectl
- pods
- rotulos e anotações
- descoberta de serviços
- balanceamento de carga HTTP com ingress
- replicasets
- deployments
- daemonsets
- jobs
- configmaps e secrets
- controle de acesso baseado em perfis no kubernetes
- integrando soluções de armazenagem no kubernetes
- estendendo o kubernetes
- implantando aplicações no mundo real
- organizando a sua aplicação

learning from data a short course

- the learning problem
- training versus testing
- the linear model
- overfitting
- three learning principles

lógica para ciência da computação

- a linguagem da logica proposicional
- a semantica da logica proposicional
- propriedades semanticas da logica proposicional
- metodos para determinação da validade de formulas da logica proposicional
- o principio da indução finita na logica
- relações semanticas entre os conectivos da logica proposicional
- um sistema axiomatico e um sistema de dedução natural na logica proposicional
- tableaux semanticos e resolução na logica proposicional
- a linguagem da logica de predicados
- a semantica da logica de predicados
- propriedades semanticas da logica de predicados
- um sistema axiomatico e um sistema de dedução natural na logica de predicados
- tableaux semanticos e resolução na logica de predicados
- resolução na logica de predicados
- programação logica

machine learning

- introduction
- concept learning and the general-to-specific ordering
- decision and learning
- artificial neural networks
- evaluating hypotheses
- bayesian learning
- computational learning theory
- instance-based learning
- genetic algorithms
- learning sets of rules
- analytical learning
- combining inductive and analytical learning
- reinforcement Learning

main memory database systems

- introduction
- history and trends
- issues and architectural choices
- systems
- active research and future directions

massively parallel databases and mapreduce systems

- introduction
- classic parallel database systems
- columnar database systems
- mapreduce systems
- dataflow systems
- conclusions

mastering data warehouse design: relational and dimensional techniques

- introduction
- fundamental relational concepts
- Understanding the Business Model
- Developing the Model
- Creating and Maintaining Keys
- Modeling the Calendar
- Modeling Hierarchies
- Modeling Transactions
- Data Warehouse Optimization
- Accommodating Business Change
- Maintaining the Models
- Deploying the Relational Solution
- Comparison of Data Warehouse Methodologies

mining of massive datasets

- Data Mining
- MapReduce and the new software stack
- Finding Similar items
- Mining Data Stream
- Link analysis
- Frequent Itemsets
- Clustering
- Advertising on the web
- Recommendation System
- Mining Social-Network Graphs
- Dimensionality Reduction
- Large-Scale Machine Learning

modern information retrieval: the concepts and technology behind search

- introduction
- modeling
- retrieval evaluation
- query languages
- query operations
- text and multimedia languages and properties
- text operations
- indexing and searching
- parallel and distributed IR
- user interfaces and visualization
- multimedia IR: models and languages
- multimedia IR: indexing and searching
- searching the web
- libraries and bibliographical systems
- digital libraries

neural networks and learning machines

- introduction
- rosenblatt's perceptron
- model building through regression
- the least-mean-square algorithm
- multiplay perceptrons
- kernel methods and radial-bases function networks
- support Vector Machines
- regularization theory
- principal-components analysis
- self-organizing maps
- information-theoretic learning models
- stochastic methods rooted in statistical mechanics
- dynamic programming
- neurodynamics
- bayseian filtering for state estimation of dynamic systems
- dynamiclly driven recurrent networks

next generation of data-mining applications

- Trends in DataMining Applications: From Research Labs to Fortune 500 Companies
- Industrial Applications
- <https://ieeexplore.ieee.org/document/5769714>
- Damage Detection Employing DataMining Techniques
- Data Projection Techniques and Their Application in Sensor Array Data Processing
- Business Applications
- <https://ieeexplore.ieee.org/document/5769709>
- Sales Opportunity Miner: Data Mining for Automatic Evaluation of Sales Opportunity
- A Fully Distributed Framework for CostSensitive Data Mining
- Application of Variable Precision Rough Set Approach to Car Driver Assessment
- Science and Engineering Applications
- Discovery of Patterns in Earth Science Data Using Data Mining
- An Active Learning Approach to Egeria Densa Detection in Digital Imagery
- Experiences in Mining Data from Computer Simulations
- Statistical Modeling of LargeScale Scientific Simulation Data
- Bioinformatics and Biotechnology Applications
- Data Mining for Gene Mapping
- DataMining Techniques for Microarray Data Analysis
- The Use of Emerging Patterns in the Analysis of Gene Expression Profiles for the Diagnosis and Understanding of Diseases
- Proteomic Data Analysis: Pattern Recognition for Medical Diagnosis and Biomarker Discovery
- Medical and Pharmaceutical Applications
- <https://ieeexplore.ieee.org/document/5769722>
- <https://ieeexplore.ieee.org/document/5769730>

operadores do kubernetes: automatizando a plataforma de orquestração de contêineres

- operadores ensinam novos truques ao kubernetes
- Executando operadores
- Operadores na interface do Kubernetes
- Operator framework
- Aplicação de exemplo: o visitors site
- Operadores Adaptadores
- Operadores em Go com o Operator SDK
- Operator Lifecycle Manager
- Filosofia dos Operadores
- Colabore
- Apendice A: Executando um Operador como um deployment em um cluster
- Apendice B: Validação de recursos personalizados
- Apendice C: Role-Based Access Control (RBAC)

pattern classification

- Bayesian Decision Theory.
- Maximum-Likelihood and Bayesian Parameter Estimation.
- Nonparametric Techniques.
- Linear Discriminant Functions.
- Multilayer Neural Networks.
- Stochastic Methods.
- Nonmetric Methods.
- Algorithm-Independent Machine Learning.
- Unsupervised Learning and Clustering.

pattern recognition and machine learning

- Introduction
- Probability Distributions
- Linear Models for Regression
- Linear Models for Classification
- Neural Networks
- Kernel Methods
- Sparse Kernel Machines
- Graphical Models
- Mixture Models and EM
- Approximate Inference
- Sampling Methods
- Continuous Latent Variables
- Sequential Data
- Combining Models

postgresql - guia do programador

- Bem-vindo ao PostgreSQL
- Instalando o PostgreSQL
- Linguagem SQL
- Gerenciando um Banco de Dados
- phpPgAdmin
- Segurança e Direitos de acesso
- Backups e restores
- Funções, Transações e Triggers
- Otimização (Tuning)
- PHP
- ASP
- Java/JSP
- -.NET

practical data science with r

- the data science process
- starting with R and data
- exploring data
- managing data
- data engineering and data shaping
- choosing and evaluating models
- linear and logistic regression
- advanced data preparation
- unsupervised methods
- exploring advanced methods
- documentation and deployment
- producing effective representation

principios de sistemas de bancos de dados distribuidos

- INTRODUÇÃO
- REQUISITOS FUNCIONAIS DE UM SGBD DISTRIBUÍDO
- ESPECIFICAÇÃO DAS INTERFACES DE UM SGBD DISTRIBUÍDO
- ESPECIFICAÇÃO DAS INTERFACES DE UM SGBD DISTRIBUÍDO
- Principais Funções de um SGBD Distribuído
- ESPECIFICAÇÃO DE BANCOS DE DADOS DISTRIBUÍDOS
- PROJETO DE BANCOS DE DADOS DISTRIBUÍDOS
- ADMINISTRAÇÃO DE BANCOS DE DADOS DISTRIBUÍDOS
- ETAPAS DO PROCESSAMENTO DE COMANDOS DA LMD
- UMA LINGUAGEM DE DEFINICAO E MANIPULAÇÃO DE DADOS RELACIONAL
- O SUBSISTEMA DE ARMAZENAMENTO
- EXEMPLO DO PROCESSAMENTO DE UMA CONSULTA
- INTRODUÇÃO AO PROCESSAMENTO DE CONSULTAS
- CLASSIFICAÇÃO DE CONSULTAS
- TRADUÇÃO PARA O ESQUEMA CONCEITUAL
- TRADUÇÃO PARA O ESQUEMA INTERNO
- OTIMIZAÇÃO
- PROCESSAMENTO DE ATUALIZAÇÕES
- INTRODUÇÃO AO PROCESSAMENTO DE CONSULTAS
- PROCESSAMENTO CENTRALIZADO COM REDUÇÃO
- DESMEMBRAMENTO CONTROLADO DISTRIBUÍDO - PARTE I
- DESMEMBRAMENTO CONTROLADO DISTRIBUÍDO - PARTE II
- PROCESSAMENTO DE ATUALIZAÇÕES
- O CONCEITO DE TRANSAÇÃO

principles of big data: preparing, sharing, and analyzing complex information

- providing structure to unstructured data
- identification, deidentification, and reidentification
- ontology and semantics
- instrospection
- data integration and software interoperability
- immutability and immortality
- measurement
- simple but powerful big data techniques
- analysis
- special considerations in big data analysis
- stepwise approach to big data analysis
- failure
- legalities
- societal issues
- the future

principles of data mining

- introduction
- measurement and data
- visualizing and exploring data
- data analysis and uncertainty
- a systematic overview of data mining algorithms
- models and patterns
- score functions for data mining algorithms
- search and optimization methods
- descriptive modeling
- predictive modeling for classification
- predictive modeling for regression
- data organization and databases
- finding patterns and rules
- retrieval by content

principles of data science

- How to Sound Like a Data Scientist
- Types of Data
- The Five Steps of Data Science
- Basic Mathematics
- Impossible or Improbable – A Gentle Introduction to Probability
- Advanced Probability
- Basic Statistics
- Advanced Statistics
- Communicating Data
- How to Tell If Your Toaster Is Learning – Machine Learning Essentials
- Predictions Don't Grow on Trees – or Do They?
- Beyond the Essentials
- Case Studies

principles of database and knowledge-base systems

- databases, object bases, and knowledge bases
- data models for database systems
- logic as a data model
- relational query language
- object-oriented databases languages
- physical data organization
- desing theory for relational databases
- protecting the database against misuse
- transaction management
- distributed database management

projeto e modelagem de banco de dados

- introdução
- o modelo entidade-relacionamento
- A UML (unified modeling language)
- analise de requisito e modelagem de dados conceitual
- transformando o modelo de dados conceitual em SQL
- normalização
- exemplo de projeto logico de banco de dados
- projeto objeto-relacional
- XML e banco de dados da web
- business intelligence
- ferramentas case para projeto logico de banco de dados

readings in database systems

- Data Models and DBMS Architecture
- Query Processing
- Data Storage and Access Methods
- Transaction Management
- Extensibility
- Database Evolution
- Data Warehousing
- Data Mining
- Web Services and Data Bases
- Stream-Based Data Management

searching multimedia databases by content

- <https://link.springer.com/chapter/10.1007/978-1-4613-1445-5_1>
- <https://link.springer.com/chapter/10.1007/978-1-4613-1445-5_2>
- <https://link.springer.com/chapter/10.1007/978-1-4613-1445-5_3>
- <https://link.springer.com/chapter/10.1007/978-1-4613-1445-5_4>
- <https://link.springer.com/chapter/10.1007/978-1-4613-1445-5_5>
- <https://link.springer.com/chapter/10.1007/978-1-4613-1445-5_6>
- <https://link.springer.com/chapter/10.1007/978-1-4613-1445-5_7>
- <https://link.springer.com/chapter/10.1007/978-1-4613-1445-5_8>
- <https://link.springer.com/chapter/10.1007/978-1-4613-1445-5_9>
- <https://link.springer.com/chapter/10.1007/978-1-4613-1445-5_10>
- <https://link.springer.com/chapter/10.1007/978-1-4613-1445-5_11>
- <https://link.springer.com/chapter/10.1007/978-1-4613-1445-5_12>
- <https://link.springer.com/chapter/10.1007/978-1-4613-1445-5_13>
- <https://link.springer.com/chapter/10.1007/978-1-4613-1445-5_14>
- <https://link.springer.com/chapter/10.1007/978-1-4613-1445-5_15>
- <https://link.springer.com/chapter/10.1007/978-1-4613-1445-5_16>

similarity search: the metric space approach (advances in database systems, 32)

- <https://link.springer.com/chapter/10.1007/0-387-29151-2_1>
- <https://link.springer.com/chapter/10.1007/0-387-29151-2_2>
- <https://link.springer.com/chapter/10.1007/0-387-29151-2_3>
- <https://link.springer.com/chapter/10.1007/0-387-29151-2_4>
- <https://link.springer.com/chapter/10.1007/0-387-29151-2_5>

sistemas de banco de dados: design, implementação e administração

- sistemas de bancos de dados
- modelos de dados
- modelo de banco de dados relacional
- modelagem entidade-relacionamento (ER)
- normalização das tabelas de banco de dados
- modelagem de dados avançada
- introdução a linguagem SQL
- sql avançada
- projeto de bancos de dados
- gerenciamento de transações e controle de concorrencia
- Sintonização (tuning) de desempenho de banco de dados e otimização de consultas
- sistemas de gerenciamento de banco de dados distribuidos
- bussiness inteligence e data warehouse
- conectividade de banco de dados e tecnologias da web
- administração e segurança de bancos de dados

sistemas de bancos de dados

- introdução
- modelo relacional
- SQL
- SQL avançada
- outras linguagens relacionais
- projeto de banco de dados e o modelo ER
- projeto de banco de dados relacional
- projeto de desenvolvimento de aplicação
- bancos de dados baseados em objeto
- XML
- armazenamento e estrutura de arquivos
- indexação e hashing
- processamento da consulta
- otimização da consulta
- transações
- controle de concorrencia
- Sistemas de recuperação
- mineração e analise de dados
- recuperação de informações
- arquiteturas de sistemas de bancos de dados
- bancos de dados paralelos
- bancos de dados distribuidos
- desenvolvimento avançado de aplicações
- tipos de dados avançados e novas aplicações
- processamento avançado de transações

sistemas de gerenciamento de banco de dados

- visão geral sobre sistemas de bancos de dados
- introdução ao projeto de banco de dados
- o modelo relacional
- algebra e calculo relacionais
- sql: consultas, restrições, gatilhos
- desenvolvimento de aplicativo de banco de dados
- aplicativos internet
- visão geral de armazenamento e indexação
- armazenando dados: discos e arquivos
- indexação estruturada em arvore
- indexação baseada em hash
- visão geral da avaliação de consultas
- ordenação externa
- avaliando operadores relacionais
- um otimizador de consultas relacional tipico
- visao geral do gerenciamento de transações
- controle de concorrencia
- recuperação de falhas
- refinamento, sintonização e formas normais
- projeto fisico de banco de dados e sintonização
- segurança e autorização
- bancos de dados paralelos e distribuidos
- sistemas de banco de dados de objetos
- bancos de dados dedutivos
- data warehousing e apoio a decisão

streaming systems: the what, where, when, and how of large-scale data processing

- streaming 101
- the what, where, when, and how of data processing
- watermarks
- advanced windowing
- exactly-once and side effects
- streams and tables
- the practicalities of persistent state
- streaming sql
- streaming joins
- the evolution of large-scale data processing

the data science design manual

- What is Data Science?
- Mathematical Preliminaries
- Data Munging
- Scores and Rankings
- Statistical Analysis
- Visualizing Data
- Mathematical Models
- Linear Algebra
- Linear and Logistic Regression
- Distance and Network Methods
- Machine Learning
- Big Data: Achieving Scale
- Coda

the data science handbook

- Introduction: Becoming a Unicorn
- The Data Science Road Map
- Programming Languages
- Data Munging: String Manipulation, Regular Expressions, and Data Cleaning
- Visualizations and Simple Metrics
- Machine Learning Overview
- Interlude: Feature Extraction Ideas
- Machine Learning Classification
- Technical Communication and Documentation
- Unsupervised Learning: Clustering and Dimensionality Reduction
- Regression
- Data Encodings and File Formats
- Big Data
- Databases
- Software Engineering Best Practices
- Natural Language Processing
- Time Series Analysis
- Probability
- Statistics
- Programming Language Concepts
- Performance and Computer Memory
- Computer Memory and Data Structures
- Maximum Likelihood Estimation and Optimization
- Advanced Classifiers
- Stochastic Modeling

the data warehouse etl toolkit: practical techniques for extracting, cleaning, conforming, and delivering data

- Introduction
- surrounding the requirements
- ETL data structures
- extracting
- cleaning and conforming
- delivering dimensions tables
- delivering fact tables
- development
- operations
- metadata
- responsabilities
- real-time ETL systems
- conclusions

the data warehouse lifecycle toolkit

- introduction
- introducing the kimball lifecycle
- launching and managing the project/program
- collecting the requirements
- introducing the technical architecture
- creating the architecture plan and selecting products
- introducing dimensional modeling
- designing the dimensional modeling
- designing the physical database and planning for performance
- introducing extract, transformation, and Load
- designing and developing the ETL system
- introducing business intelligence applications
- designing and developing business intelligence applications
- deploying and supporting the DW/BI System
- expanding the DW/BI system

the data warehouse toolkit: the complete guide to dimensional modeling

- data warehousing, business intelligence, and dimensional modeling primer
- kimball dimensional modeling techniques overview
- retail sales
- inventory
- Procurement
- order management
- accounting
- customer relationship management
- human resource management
- financial services
- telecommunications
- transportation
- education
- healthcare
- electronic commerce
- insurance
- kimball lifecycle overview
- dimensional modeling process and tasks
- ETL subsystems and techniques
- ETL system design and development process and tasks
- big data analytics

the design and implementation of modern column-oriented database systems (foundations and trends(r) in databases)

- introduction
- history, trends and performance tradeoffs
- column-store architectures
- column-store internals and advanced techniques
- Discussions, conclusions and future directions

the elements of statistical learning: data mining, inference, and prediction

- Introduction
- Overview of Supervised Learning
- Linear Methods for Regression
- Linear Methods for Classification
- Basis Expansions and Regularization
- Kernel Smoothing Methods
- Model Assessment and Selection
- Model Inference and Averaging
- Additive Models, Trees, and Related Methods
- Boosting and Additive Trees
- Neural Networks
- Support Vector Machines and Flexible Discriminants
- Prototype Methods and Nearest-Neighbors
- Unsupervised Learning
- Random Forests
- Ensemble Learning
- Undirected Graphical Models
- High-Dimensional Problems: p -> N

the kubernetes book

- kubernetes primer
- kubernetes principles of operation
- getting kubernetes
- working with pods
- virtual clusters with namespaces
- kubernetes deployments
- kubernetes services
- ingress
- service discovery deep dive
- kubernetes storage
- configmaps and secrets
- statefulsets
- API secutiry and RBAC
- The Kubernetes API
- Threat modeling kubernetes
- real-world kubernetes security

truthful art, the: data, charts, and maps for communication

- introduction
- what we talk about when we talk about visualization
- the five qualitites of great visualizations
- the truth continuum
- of conjectures and uncertainty
- basic principles of visualization
- exploring data with simple charts
- visualizing distributions
- revealing change
- seeing relationships
- mapping data
- uncertainty and significance
- on creativity and innovation

utilizando uml e padrões: uma introdução à análise e ao projeto orientados a objetos ao desenvolvimento iterativo

- Análise e Projeto Orientados a Objetos
- Iterativo, Evolutivo e Agil
- Estudos de Caso
- Concepção não é a Fase de Requisitos
- Requisitos Evolutivos
- Casos de Uso
- Outros Requisitos
- Iteração 1 - Conceitos Básicos
- Modelos de Domínio
- Diagramas de Sequência do Sistema
- Contratos de Operação
- Dos Requisitos para o Projeto - Iterativamente
- Arquitetura Lógica e Diagrama de Pacotes UML
- Projeto de Objetos
- Diagramas de Interação UML
- Diagramas de Classe UML
- GRASP: Projeto de Objetos com Responsabilidades
- Exemplos de Projeto de Objetos com GRASP
- Projetar para Visibilidade
- Mapeamento de Projetos para Código
- Desenvolvimento Dirigido por Teste e Refatoração
- Ferramentas UML e UML como Documentação
- Rápida Atualização da Análise
- Iteração 2 - Requisitos
- GRASP: Mais Objetos com Responsabilidades

web database applications with php & mysql, 2nd edition

- Database Applications and the Web
- The PHP Scripting Language
- Arrays, Strings, and Advanced Data Manipulation in PHP
- Introduction to Object-Oriented Programming with PHP 5
- SQL and MySQL
- Querying Web Databases
- PEAR
- Writing to Web Databases
- Validation with PHP and JavaScript
- Sessions
- Authentication and Security
- Errors, Debugging, and Deployment
- Reporting
- Advanced Features of Object-Oriented Programming in PHP 5
- Advanced SQL
- Hugh and Dave's Online Wines:A Case Study
- Managing Customers
- The Shopping Cart
- Ordering and Shipping at the Online Winestore
- Searching and Authentication in the Online Winestore

F

TAXONOMIA DA BIBLIOGRAFIA

Abaixo você encontra a definição da Taxonomia da Bibliografia, gerada a partir dos resultados da nossa pesquisa.

Tabela 8 – Tópicos de Bancos de Dados e Computação

1º Nível	**2º Nível**	**3º Nível**
Bancos de dados relacionais	Modelo relacional	
	Projeto de banco de dados relacionais	
	Algebra e Calculo Relacional	
	SQL	SQL embutida
		SQL dinâmica
		Consulta Recursiva
		Restrições e Gatilhos
		Procedimentos Armazenados
		Visões
	Forma Normal	
	Integridade de Dados	
	Calculo de predicados	
	Interface Web com banco de dados relacionais	PHP
		Autenticação e Segurança
	Sessão	
	MySQL	
	Postgres	
	Encoding	

1º Nível	2º Nível	3º Nível
	Banco de dados memória principal	
	Aplicações em Bancos de Dados relacionais	
Modelagem de Dados	Modelo Entidade-Relacionamento	Mapeamento de esquema do ER para Relacional
		Mapeamento de operações do ER para Relacional
		Engenharia reversa de esquema do relacional para ER
	Conceitos de Modelagem de Dados	Estratégias de modelagem de Esquema
		Qualidade de dados
		Análise funcional usando Data Flow
	Regras de Negócio	
	Network Model	
	Modelo Entidade-Relacionamento Estendido	
	UML	
	Projeto usando mapeamento ER e EER	
Arquitetura de bancos de dados	cliente-servidor	bulk loading
	processamento distribuído	
	centralizada	
Segurança em bancos de dados	Controle de acesso baseado em concessão e revogação de privilégios	

1º Nível	2º Nível	3º Nível
	Controle de acesso obrigatório e controle de acesso baseado em papel para segurança multinível	
	Injeção de SQL	
	Criptografia e infraestrutura de chave pública	
	Banco de Dados Estatístico	
Modelo Objeto-Relacional	Operações sobre Objeto-Relacional	
Data Warehouse	OLAP	
	Cubo de dados	
Dado Semiestruturados	XML	
Projeto Físico de banco de dados relacionais	Técnicas de Armazenamento	Clusterização
	Indexação	Dependências Funcionais
		Sparse vs Dense Index
		Índice Distribuído
		Acesso por chave primária
		Acesso por chave secundária
		Indexação de imagem
	Hashing	
	Otimização de SO	
	Tunning de Hardware	
Processamento e otimização de consultas	Plano de execução	
	Algoritmo de execução de consultas	
	Gerenciamento de Buffer	
	Compilação de consultas	
	Transação	
	Recuperação	
Recuperação de falhas	Logging	

1º Nível	2º Nível	3º Nível
Controle de concorrência	Escalonamento	
	Serializabilidade	
	Protocolos de bloqueios	
	Métodos Otimistas	
	Timestamp	
	Deadlock	
Banco de dados Distribuídos	Processamento e Otimização de Consulta	Profiling de consulta
	Arquitetura	
	Comprometimento em duas fases	
	Bloqueio distribuído	
	Transações distribuídas	
	Controle de Concorrência	
	Replicação de Dados	
	Gerência de Transações	
	Controle de integridade	
	LMD	
	Administração de Banco de Dados Distribuídos	
	Segurança	
Banco de dados Paralelos	MapReduce	
	Processamento consulta paralelo	
Mineração de dados	Mineração de items-frequentes	
	Mineração de items similares	
	Clusterização	Análise de cluster
	Streaming de dados	
	Data preprocessing	
	Data Cleaning	
	Data Integration	
	Data Reduction	

1º Nível	2º Nível	3º Nível
	Data Transformation	
	Detecção de outlier	
	Missing Values	
	Visualização de dados	
	Análise de dados	
	Algoritmos de data mining	
		Modelagem para predição
		Distribuição de probabilidade
		Função de Score
	KDD	
	Mineração de Grafos	
	Modelagem estatística	
Recuperação de informação	Sistema de buscas	Metric Space Searching
		Centralized Index Structures
		Approximate Similarity Search
		Parallel e Distributed Index
	PageRank	
	Lista Invertidas	
	Modelo Vetorial	
	Word Embeddings	
	Indexação e Busca	
	IR paralelo e distribuído	
	Visualização	
	Multimedia IR	
	Operação em Texto	
	Query	
	Métricas de Recuperação	
	Modelos de Recuperação de Informação	
	Content-Based Image and Video Retrieval	
Business Intelligence	Arquitetura	

1º Nível	2º Nível	3º Nível
Gerenciamento de Bancos de dados relacionais	Instalação e Upgrade	
	Padrões e Procedimentos	
	Design de Bancos de Dados Relacionais	
	Denormalização	
	Performance e Tuning	
	Gerenciamento de Mudança	
	Disponibilidade de dados	
	Planejamento de Desastre	
	Gerenciamento de armazenamento	
	Gerenciamento de Metadados	
	Garantia de Nível de Serviço	
	Segurança	
	Gerenciamento de Infraestrutura	
	Gerenciamento de Risco	
Banco de dados dedutivos	modelo mínimo	
	consultas recursivas com negação	
	DataLog	
	DataLog para SQL	
gerenciamento de dados espaciais	tipos de dados	
	consultas	
	índices espaciais	
	Indexação Baseada em Curvas de Preenchimento de Espaço	
	Arquivos Grid	
	Árvores R	
Machine Learning	Classificação	

1º Nível	2º Nível	3º Nível
	Regressão	
	Árvore de Decisão	
	Cross Validation	
	Rede Bayesiana	
Estrutura de Dados multidimensional	Multidimensional Point Data	
	Object-Based e Image-based image representation	
	interval and small Rectangle	
	High-Dimensional Data	
Cloud Computing	Tipos de Cloud	
	Gerenciamento de infraestrutura	
	Arquitetura de Cloud Computing	
	Infraestrutura como Serviço	
	Migração para Cloud	
	Integração de Saas	
	Enterprise Cloud	Cloud Supply Chain
	Virtualização	
	Segurança	
	Aplicações em Cloud Computing	MapReduce
		Aplicações científicas
		Workflow Engine
	Monitoração	
	Gerenciamento de SLA	
	Arquitetura de aplicações em Cloud	
	Gerenciamento de mudança	
	Escalonamento	
	Containerization	Kubernetes
		Operadores de Kubernetes

1º Nível	2º Nível	3º Nível
		Docker
	Mecanismos de	
	Infraestrutura de Cloud	
	Cloud Delivery Model	
	Pricing Model e Cost Metrics	
Sistemas Distribuídos	Network-Based Systems	
	Ecossistema de software	
	Modelos de sistemas	
	Clusterização para	
	paralelismo massivo	
	Arquitetura orientada a	
	serviço	
	Computação em grid	
	Computação em P2P	
	Computação Ubíqua	
	Micro-Serviços	Arquitetura
		Comunicação entre
		micro-serviços
		Workflow
		Implantação
		Testes
		Observabilidade
		Segurança
		Resiliência
		Escalabilidade
	Single-Node Pattern	Sidecar
		Ambassador
		Adapters
	Serving Pattern	Replicated Load-Balanced
		Services
		Sharded Services
		Scatter/Gather
		Event-Driven Processing

1º Nível	2º Nível	3º Nível
		Ownership Election
	Batch Pattern	Work Queue Systems
		Event-Driven Batch Processing
		Coordinated Batch Processing
	Consistência e Consenso	
Data Streaming	Marca D'água	
	Windowing	
	Exactly-Once	
	SQL	
	Arquitetura de Processamento de Dados	Kafka
Column-Oriented Database Systems		
Big Data	Dado estruturado	
	Dado não-estruturado	
	Identificação	
	Ontologia	
	Integração de dados	
	Imutabilidade	
	Análise de dados	
	Legal	
	Falha	
	Arquiteturas	Data-Driven
	Integração Big Data para DW	Gerenciamento de ciclo de vida de dados
		Visualização de dados
	Business Intelligence em Big Data	
SRE	Gerenciamento de infraestrutura	
	Riscos	
	Nível de Serviço	

1º Nível	**2º Nível**	**3º Nível**
	Monitoração	
	Engenharia de release	
	Oncall	
	Testes	
	Load Balancer	
	Falhas	
	Escalonamento	
	Pipeline de processamento de dados	
	Integridade de dados	

G

Taxonomia das Ementas

Abaixo você encontra a definição da Taxonomia das Ementas, com frequência, gerada a partir dos resultados da nossa pesquisa.

Tabela 9 – Tópicos e Frequências dos Capítulos

1º Nível	**2º Nível**	**3º Nível**
Modelo de Dados pré-relacionais (2)	Modelos de Dados: Hierárquico, Redes (4)	
Bancos de dados não relacionais (5)		
Bancos de dados não convencionais (3)	bancos de dados heterogêneos (3) Bancos de dados de genoma (1) BD TEMPORAIS, SISTEMAS DE INFORMAÇÕES GEOGRÁFICAS, BD MULTIMÍDIA, ABORDAGEM TOP-DOWN, ABORDAGEM BOTTOM-UP, BD HOMOGÊNEO X HETEROGÊNEO, FRAGMENTAÇÃO, REPLICAÇÃO E ALOCAÇÃO (3)	

1º Nível	2º Nível	3º Nível
Projeto Físico de banco de dados relacionais (16)	Técnicas de Armazenamento (15)	
	Indexação (14)	
	Hashing (2)	
	Tunning de Hardware (7)	
	BANCO DE DADOS AUTONÔMOS (1)	
	BANCO DE DADOS AUTO-CONFIGURÁVEIS (1)	
	Gasto energético e meio ambiente (1)	
Bancos de dados relacionais (20)	Aplicações em Bancos de Dados relacionais (7)	
	Modelo Relacional (16)	
	SQL (21)	Visoes (4)
		Restrições e Gatilhos (15)
		Procedimentos Armazenados (5)
	Algebra Relacional (13)	
	Forma Normal (16)	
	Interface Web com banco de dados relacionais (4)	
	Integridade de Dados (2)	
Processamento e otimização de consultas (18)	previsão de seletividade (3)	
Controle de concorrencia (20)	Deadlock (3)	
	Metodos Otimistas (2)	
	Timestamp (1)	
	Transação (20)	
	Protocolos de bloqueios (1)	
Recuperação de falhas (11)	Logging (3)	

1º Nível	2º Nível	3º Nível
	Recuperação (13)	Recuperação de falhas: categorias de falhas, gerência de buffer, técnicas de recuperação (3) Cópia e recuperação da base de dados (1)
Modelagem Conceitual (14)	Modelo Entidade-Relacionamento (15)	Mapeamento de esquema do ER para Relacional (7) Mapeamento de operações do ER para Relacional (6)
	Modelo Entidade-Relacionamento Estendido (3) UML (4)	
Dado Semiestruturados (2)	XML (2)	
Mineração de dados (8)	Clusterização (2) Streaming de dados (1) Data preprocessing (1) Data Cleaning (1) Data Integration (1) Data Transformation (1) Detecção de outlier (1) Missing Values (1) Visualização de dados (1) Algoritmos de data mining (2) KDD (2) mineração de dados em domínios específicos (2)	Analise de cluster (1)
Data Warehouse (6)	OLAP (4) Cubo de dados (1)	

1º Nível	2º Nível	3º Nível
Modelo Objeto-Relacional (10)	Operações sobre Objeto-Relacional (2)	
Arquitetura de bancos de dados (10)	cliente-servidor (2)	
	bancos de dados em ambientes de computação móvel (2)	
	ANSI-X3-SPARC (1)	
Segurança em bancos de dados (11)	Controle de acesso baseado em concessão e revogação de privilegios (3)	
	Controle de acesso obrigatório e controle de acesso baseado em papel para segurança multinível (1)	
	Injeção de SQL	
	Criptografia e infraestrutura de chave publica	
Gerenciamento de Bancos de dados relacionais (16)	Instalação e Upgrade (2)	
	Padrões e Procedimentos (2)	
	Design de Bancos de Dados Relacionais (2)	
	Denormalização	
	Performance e Tuning (8)	
Banco de dados Distribuidos (8)	Arquiteturas (4)	
	Controle de Concorrencia (2)	
	Administração de Banco de Dados Distribuidos (3)	
	Processamento e Otimização de Consulta (4)	

1º Nível	2º Nível	3º Nível
	Particionamento e alocação	
	de dados e aplicações (2)	
	Bloqueio distribuido (1)	
	Transações distribuidas (1)	
	Replicação de Dados (3)	
	Gerencia de Transações (3)	
	Controle de integridade (1)	
Business Inteligence (3)		
gerenciamento de dados		
espaciais (1)		
Banco de dados Paralelos (1)	Processamento consulta	
	paralelo (1)	
Machine Learning	Classificação (1)	
	Regressão (1)	
Recupeção de informação (5)	Sistema de buscas (2)	

H

Dados brutos da pesquisa Habilidades para Engenheiros de Dados

Os nomes e as empresas dos respondentes foram anonimizados para não identificação dos mesmo.

Tabela 10 – Pesquisa junto a industria brasileira de tecnologia

Perguntas e Respostas
Cargo atual

1. Senior Data Scientist
2. Gerente de Produto
3. Engenheiro de Dados
4. Diretor de Engenharia
5. Engenheiro de Dados Sênior
6. Diretor de Tecnologia
7. Head de Dados
8. Tribe Lead
9. CTO
10. gerente

Nível de conhecimento em cada habilidade de um Engenheiro de Dados no inicio da carreira (2 anos ou menos de experiencia) [SQL]

1. Conhecimento pratico avançado
2. Conhecimento pratico basico
3. Conhecimento pratico basico
4. Conhecimento pratico basico
5. Conhecimento teorico
6. Conhecimento teorico
7. Conhecimento pratico avançado
8. Conhecimento pratico basico
9. Conhecimento pratico basico
10. Conhecimento pratico basico

Nível de conhecimento em cada habilidade de um Engenheiro de Dados no inicio da carreira (2 anos ou menos de experiencia) [Modelagem de dados relacionais]

1. Conhecimento pratico basico
2. Conhecimento pratico basico
3. Conhecimento pratico basico
4. Conhecimento pratico basico
5. Conhecimento teorico
6. Conhecimento pratico basico
7. Conhecimento teorico
8. Conhecimento pratico basico
9. Conhecimento pratico basico
10. Conhecimento teorico

Nível de conhecimento em cada habilidade de um Engenheiro de Dados no inicio da carreira (2 anos ou menos de experiencia) [NoSQL databases]

1. Conhecimento teorico
2. Conhecimento pratico basico
3. Conhecimento pratico avançado
4. Conhecimento pratico basico
5. Conhecimento teorico
6. Conhecimento teorico
7. Conhecimento teorico
8. Conhecimento teorico
9. Conhecimento teorico
10. Conhecimento teorico

Nível de conhecimento em cada habilidade de um Engenheiro de Dados no inicio da carreira (2 anos ou menos de experiencia) [Criação e gerenciamento de ETLs]

1. Conhecimento pratico basico
2. Conhecimento pratico basico
3. Conhecimento pratico basico
4. Conhecimento pratico avançado
5. Conhecimento pratico basico
6. Conhecimento pratico basico
7. Conhecimento pratico basico
8. Conhecimento pratico basico
9. Conhecimento pratico basico
10. Conhecimento pratico avançado

Nível de conhecimento em cada habilidade de um Engenheiro de Dados no inicio da carreira (2 anos ou menos de experiencia) [Streaming processing]

1. Conhecimento teorico
2. Conhecimento teorico
3. Conhecimento pratico avançado
4. Conhecimento pratico basico
5. Conhecimento pratico basico
6. Conhecimento pratico basico
7. Conhecimento pratico basico
8. Conhecimento teorico
9. Conhecimento teorico
10. Conhecimento teorico

Nível de conhecimento em cada habilidade de um Engenheiro de Dados no inicio da carreira (2 anos ou menos de experiencia) [Batch processing]

1. Conhecimento pratico basico
2. Conhecimento pratico basico
3. Conhecimento pratico avançado
4. Conhecimento pratico basico
5. Conhecimento pratico basico
6. Conhecimento pratico basico
7. Conhecimento pratico basico
8. Conhecimento pratico basico
9. Conhecimento teorico
10. Conhecimento pratico basico

Nível de conhecimento em cada habilidade de um Engenheiro de Dados no inicio da carreira (2 anos ou menos de experiencia) [Distributed computing]

1. Conhecimento pratico basico
2. Conhecimento pratico basico
3. Conhecimento pratico basico
4. Conhecimento pratico avançado
5. Conhecimento teorico
6. Conhecimento teorico
7. Conhecimento teorico
8. Conhecimento teorico
9. Conhecimento teorico
10. Conhecimento teorico

Nível de conhecimento em cada habilidade de um Engenheiro de Dados no inicio da carreira (2 anos ou menos de experiencia) [Distributed file system]

1. Conhecimento teorico
2. Conhecimento pratico basico
3. Conhecimento pratico avançado
4. Conhecimento pratico avançado
5. Conhecimento teorico
6. Conhecimento teorico
7. Conhecimento teorico
8. Conhecimento teorico
9. Conhecimento teorico
10. Conhecimento teorico

Nível de conhecimento em cada habilidade de um Engenheiro de Dados no inicio da carreira (2 anos ou menos de experiencia) [Cloud computing]

1. Conhecimento pratico basico
2. Conhecimento pratico basico
3. Conhecimento pratico basico
4. Conhecimento pratico basico
5. Conhecimento teorico
6. Conhecimento pratico basico
7. Conhecimento pratico basico
8. Conhecimento pratico basico
9. Conhecimento teorico
10. Conhecimento pratico basico

Nível de conhecimento em cada habilidade de um Engenheiro de Dados no inicio da carreira (2 anos ou menos de experiencia) [Arquitetura de data lake]

1. Conhecimento teorico
2. Conhecimento teorico
3. Conhecimento pratico avançado
4. Conhecimento pratico avançado
5. Conhecimento pratico basico
6. Conhecimento teorico
7. Conhecimento pratico basico
8. Conhecimento teorico
9. Conhecimento teorico
10. Conhecimento teorico

Nível de conhecimento em cada habilidade de um Engenheiro de Dados no inicio da carreira (2 anos ou menos de experiencia) [Arquitetura de data warehouse]

1. Conhecimento teorico
2. Conhecimento teorico
3. Conhecimento pratico avançado
4. Conhecimento pratico avançado
5. Conhecimento pratico basico
6. Conhecimento teorico
7. Conhecimento pratico basico
8. Conhecimento pratico basico
9. Conhecimento teorico
10. Conhecimento teorico

Nível de conhecimento em cada habilidade de um Engenheiro de Dados no inicio da carreira (2 anos ou menos de experiencia) [Arquitetura de data lakehouse]

1. Conhecimento teorico
2. Conhecimento teorico
3. Conhecimento pratico avançado
4. Conhecimento pratico avançado
5. Conhecimento teorico
6. Conhecimento teorico
7. Conhecimento pratico basico
8. Conhecimento teorico
9. Conhecimento teorico
10. Conhecimento teorico

Nível de conhecimento em cada habilidade de um Engenheiro de Dados no inicio da carreira (2 anos ou menos de experiencia) [Programação]

1. Conhecimento pratico basico
2. Conhecimento pratico avançado
3. Conhecimento pratico avançado
4. Conhecimento pratico basico
5. Conhecimento pratico avançado
6. Conhecimento pratico avançado
7. Conhecimento pratico avançado
8. Conhecimento pratico basico
9. Conhecimento pratico basico
10. Conhecimento pratico avançado

Nível de conhecimento em cada habilidade de um Engenheiro de Dados no inicio da carreira (2 anos ou menos de experiencia) [Automatização de Infraestrutura]

1. Conhecimento teorico
2. Conhecimento teorico
3. Conhecimento pratico basico
4. Conhecimento pratico basico
5. Conhecimento teorico
6. Conhecimento pratico basico
7. Conhecimento pratico basico
8. Conhecimento teorico
9. Conhecimento teorico
10. Conhecimento pratico basico

Gostaria de detalhar suas escolhas na pergunta acima?

1. Entendo que um engenheiro de dados com 2 anos ou menos de experiência não tenha adquirido relvante bagagem na academia e no mercado. Por isso, acredito que ele tenha dificuldade em construir uma arquitetura de dados, mas que já tenha trabalhado com a construção de pipelines de ETL simples, utilizando ferramentas com arquiteturas distribuídas para processamento eficiente dos dados. Além disso, é esperado que ele tenha tido contato básico com ferramentas e conceitos de computação nas nuvens e NoSQL. Em relação ao SQL é uma linguagem que geralmente se aprende na graduação e é bastante exercitado no mercado. Já programação, é esperado ter um conhecimento prático avançado, principalmente na construção de algoritmos e boas práticas de desenvolvimento de software.

2. Conhecimento prático avançado relativo e na medida do possível para dois anos de experiência profissional. Não espero que seja um conhecimento avançado no absoluto, isto requer mais tempo.

3. Direcionei minhas respostas de acordo com potenciais atividades que esse engenheiro poderia atuar ativamente. Processar dados via Stream, batchs e criação de ETL's podem se encaixar bem com o nível de profissional com 2 anos ou menos.

   Acabei respondendo o básico para arquitetura pois é necessário ter mais experiência para propor soluções de arquitetura compatíveis com os problemas à serem resolvidos. Normalmente um arquiteto deveria ter tido experiência em vários dos itens listados para conseguir desenhar uma solução de arquitetura eficiente.

   Assinalei "cloud computing"como prático porque atualmente temos muitas tecnologias e novos serviços que já nascem baseados em nuvem, é uma realidade de mercado. Do mesmo modo para "Automatização de Infraestrutura", quando se fala em Cloud, normalmente já se pensa em deixar a infraestrutura no modelo "as code"para facilitar a gestão dela.

   Além disso, as respostas estão com viés do local que trabalho atualmente. Elas podem ser diferentes de acordo com a maturidade ou necessidade de outras empresas.

4. Entendo que é uma pessoa nova de carreira, então não teria conhecimento avançado em quase nada

Habilidades não listadas acima que você considera importante para engenheiros de dados em início de carreira (2 anos ou menos de XP)

1. Inglês escrito e leitura Capacidade de articulação em termos de abstrações e noções de engenharia de software (design patterns, noção de algoritmos e complexidade, senso de qualidade de código) Inteligência emocional para ambientes profissionais
2. Princípios básicos sobre SRE, assim como aplicações ao contexto de monitoramento e confiabilidade de plataformas de dados, assim como conceitos e principios de DevOps aplicados a dados (data ops). Em resumo, times de plataforma de dados tem como objetivo garantir a disponibilidade e confiabilidade de dados, assim como garantir a autonomia de desenvolvimento técnico de qualquer outro papel analítico. Conhecimentos sobre boas práticas e principios de Site Realiabity Engineering, potencializam o atingimento do primeiro objetivo. Conhecimentos sobre data ops suportam o atingimento do segundo objetivo (diminuição do custo de desenvolvimento).

I

Taxonomia da indústria

Abaixo você encontra a definição da Taxonomia da Indústria, gerada a partir dos resultados da nossa pesquisa.

Tabela 11 – Tópicos de Bancos de Dados e Computação

1º Nível	**2º Nível**	**3º Nível**
Bancos de dados relacionais	SQL	
	Gerenciamento de bancos de dados	
	Forma Normal	
	Integridade de dados	
	Projeto Físico	Armazenamento
		Indexação
		Escalonamento
		Tunning
	Controle de Concorrência	
	Otimização de consulta	
	Bancos de dados Distribuídos	
Modelagem Conceitual	Modelo entidade-relacionamento	Mapeamento de esquema do ER para Relacional
	UML	
Bancos de dados não relacionais	Tipos de bancos	Grafos
		Documento
		Chave-Valor

1º Nível	2º Nível	3º Nível
		Tabular
	Projeto Físico	Indexação
		Armazenamento
		Escalonamento
	Integridade de dados	
	Gerenciamento de bancos de dados	
	Controle de Concorrência	
	Otimização de consulta	
Mineração de dados	Batch	
	Streaming	
	ETLs	
	Ingestão de dados	data integration
	Transformação de dados	
	Qualidade de dados	Nível sintático
		Nível semântico
		Dados faltantes
		limpeza de dados
Computação em Nuvem	Automatização de Infraestrutura	
	Técnicas de SRE	
	Containers	
	Custos	
	Virtualização	
	Armazenamento físico	
	Segurança	
	Monitoramento	
	Escalabilidade	
	Aplicações em cloud computing	
Data Warehouse	OLAP/Cubo de dados	
Big Data	Arquitetura	data lakehouse
		data lake

1º Nível	2º Nível	3º Nível
	Armazenamento Distribuído	Particionamento
		Replicação
		Transação
		Consistência e Consenso
		Falha total e parcial
	Tipos de dados	Estruturado
		Não-estruturado
	Integração de dados	
	Evolução de esquema	
	Banco de dados NewSQL	
	Governança de dados	
	MapReduce	
	Apache Spark	
Programação	Micro-serviços	
	Padrões de projeto	
	Estrutura de Dados	

J

Bibliografia nova completa

- **Título:** Descomplicando o Docker
  Ano: 2018
  Autores: Jeferson Fernando Noronha Vitalino, Marcus André Nunes Castro
  ISBN/DOI: 9788574529011
  Edição: 2

- **Título:** Pro Git
  Ano: 2014
  Autores: Chacon, Scott; Straub, Ben
  ISBN/DOI: 9781484200773
  Edição: 2

- **Título:** Google Cloud Cookbook
  Ano: 2021
  Autores: Rui Santos Costa, Drew Hodun
  ISBN/DOI: 9781492092896
  Edição: 1

- **Título:** Data Engineering with Google Cloud Platform
  Ano: 2022
  Autores: Adi Wijaya
  ISBN/DOI: 9781800561328
  Edição: 1

- **Título:** Design Cloud Data Platform
  Ano: 2021

Autores: Zburivsky, Danil; Partner, Lynda
ISBN/DOI: 9781617296444
Edição: 1

- **Título:** The Cloud Data Lake
  Ano: 2022
  Autores: Gopalan, Rukmani
  ISBN/DOI: 9781098116583
  Edição: 1

- **Título:** The Data Warehouse Toolkit
  Ano: 2013
  Autores: Kimball, Ralph; Ross, Margy; Thornthwaite, Warren; Mundy, Joy; Becker, Bob; Caserta, Joe
  ISBN/DOI: 9781118875186
  Edição: 3

- **Título:** Data Pipelines with Apache Airflow
  Ano: 2021
  Autores: Harenslak, Bas P.; de Ruiter, Julian Rutger
  ISBN/DOI: 9781617296901
  Edição: 1

- **Título:** O Grande Microlivro de Git
  Ano: 2023
  Autores: Jessica Temporal
  ISBN/DOI: -
  Edição: 1

- **Título:** Lakehouse: A New Generation of Open Platforms that Unify Data Warehousing and Advanced Analytics
  Ano: 2021
  Autores: Michael Armbrust; Ali Ghodsi; Reynold Xin; Matei Zaharia
  ISBN/DOI: -
  Edição: -

- **Título:** "Delta Lake: High-Performance ACID Table Storage over Cloud Object Stores"
  Ano: 2020

Autores: Armbrust et al.

ISBN/DOI: 10.14778/3415478.3415560

Edição: -

K

Pesquisa Qualitativa da Disciplina de Engenharia de Dados

Tabela 12 – Pesquisa Qualitativa da Disciplina de Engenharia de Dados ENGD 2024-01

Pergunta	**Escala / Resposta**
Você entende as diferentes gerações de Plataforma de Dados? (DW/Data Lake/Lakehouse)	1 (Não compreendo) – 5 (Compreendo completamente)
Você compreende as diferentes aplicações de bancos de dados Relacionais e NoSQL?	1 – 5
A disciplina abrangeu vários assuntos da ciência da computação, como Cloud Computing e Programação?	1 – 5
Você entende como e por que implementar Processamento de Dados em Lote e Streaming?	1 – 5
Descreva sua experiência quanto à parte técnica da disciplina	Resposta aberta
A disciplina foi apresentada com objetivos claros?	1 (Não concordo) – 5 (Concordo totalmente)
A disciplina é relevante para sua formação?	1 – 5
O material didático fornecido ou citado é?	1 (Muito baixo) – 5 (Muito alto)
Qual o grau de dificuldade da disciplina?	1 – 5

Continuação da Tabela 12

Pergunta	**Escala / Resposta**
Esta disciplina precisa de mais tempo semanal (mais créditos)?	1 – 5
A disciplina teve espaço para prática dos alunos (laboratório)?	1 – 5
Os trabalhos práticos refletiram o conteúdo ensinado?	1 – 5
Os pré-requisitos da disciplina são adequados?	1 – 5
Após cursar a disciplina, seu interesse pelo assunto aumentou?	1 – 5
Você adquiriu conhecimentos novos com esta disciplina?	1 – 5
Espaço para justificar suas respostas e deixar feedback sobre a disciplina	Resposta aberta
Você faz estágio ou trabalha na área atualmente?	Sim / Não
Você entende serem relevantes para o trabalho do dia a dia os conteúdos ensinados?	1 (Discordo totalmente) – 5 (Concordo totalmente)
Você entende que os conteúdos dados podem ser aplicados no dia a dia?	1 – 5
Discorra sobre como os conteúdos ensinados podem ser aplicados no dia a dia	Resposta aberta

L

Avaliação da Disciplina de Engenharia de Dados

Tabela 13 – Avaliação da Disciplina de Engenharia de Dados

Pergunta	**Escala / Resposta**
Os assuntos tratados são relevantes no dia a dia da indústria e sociedade	1 (Nada relevantes) - 5 (Completamente relevantes)
Com essas habilidades adquiridas, os alunos conseguiriam performar bem na sua empresa	1 (Não conseguiriam) - 5 (Certamente conseguiriam)
Essa disciplina consegue preparar um aluno para entrar no mercado de trabalho como Engenheiro de Dados	1 (Não é suficiente) - 5 (É totalmente suficiente)
Algum assunto relevante deveria ser tratado e não foi dado?	Resposta aberta
Comentários gerais que julgar pertinente.	Resposta aberta