

Conteúdo Programático, Bibliografia (indicação opcional) e Sistematização da Prova Prática (quando houver)

Edital UFRJ nº 54, de 30 de janeiro de 2024

Haverá Prova Prática: () Sim (X) Não

Instituto de Computação

Código da Opção de Vaga	MC-093	Departamento ou Programa / Setorização Definitiva	Computação / Engenharia de Dados
Conteúdo Programático	<ol style="list-style-type: none">1. Conceitos de Distribuição: Tipos e Arquiteturas de Bancos de Dados Distribuídos; Fragmentação, Replicação e Alocação; Sistemas de Arquivos Distribuídos2. Modelos e arquiteturas de armazenamento de dados: Armazenamento de grandes volumes de dados; Bancos de Dados NoSQL; Bancos de Dados NewSQL; Bancos de Dados em Memória3. Arquiteturas de tratamento e análise de dados: Data Warehouse; Data Lake4. Streaming de dados: Histórico e evolução; Modelos de Programação; Streaming distribuído5. Computação em nuvem: Modelos de Implantação de Nuvem - Nuvens públicas, privadas e híbridas; Modalidades de Serviços em Nuvem; Estratégias de migração; Bancos de Dados como Serviço6. Empacotamento e portabilidade de dados: Containers - construção, registro, execução e orquestração7. Modelos de tratamento de dados: Processamento de dados em lote, MapReduce e seu ecossistema; Ciclo de tratamento de dados: coleta, preparação, armazenamento, limpeza, integração, transformação; ETL x ELT8. Governança de Dados e ambientes para dados heterogêneos distribuídos: DAMA-DMBOK e Modelos de Maturidade de Gestão de Dados; Gestão de Metadados: Catálogos e Dicionários de Dados; Dados Mestres e Dados de Referência; Arquiteturas, plataformas e soluções para dados abertos9. Proteção e Privacidade de Dados em ambientes descentralizados: Proteção ao longo do ciclo de vida do dado (Privacy by Design); Lei Brasileira de Proteção de Dados; Infraestrutura de chave pública, autoridade certificadora e outros papéis; Blockchain e Tecnologia de Registro Distribuído10. Web Semântica: Dados Conectados; Representação de Dados em RDF e RDF Schema; Triplificação de Dados; Consultas SPARQL; Grafos de Conhecimento		
Bibliografia (indicação opcional)	<ol style="list-style-type: none">1. Alex Gorelik. The Enterprise Big Data Lake: Delivering the Promise of Big Data and Data Science, Editora O'Reilly, eBook Kindle (2019)2. Barbieri, Carlos. Governança de dados: práticas, conceitos e novos caminhos. Alta Books (2019)3. Daniel Donda. Guia Prático de Implementação da LGPD. Editora Labrador (2020)4. Danil Zburivsky, Lynda Partner. Designing Cloud Data Platforms. Editora Manning (2021)5. Fernando Amaral. Introdução à Ciência de Dados: Mineração de dados e big data. Alta Books (2018)6. James Hendler, Fabien Gandon, Dean Allemang. Semantic Web for the Working Ontologist: Effective Modeling for Linked		

	<p>Data, Rdfs, and Owl. Editora ACM (2020)</p> <ol style="list-style-type: none">7. Joe Reis, Matt Housley. Fundamentals of Data Engineering: Plan and Build Robust Data Systems. Editora O'Reilly (2022)8. Jule Hintzbergen, Kees Hintzbergen, André Smulders, Hans Baars. Fundamentos de Segurança da Informação: com Base na ISO 27001 e na ISO 27002. Editora Brasport (2018)9. Martin Fowler, Pramod J. Sadalage. NoSQL Essencial: um Guia Conciso Para o Mundo Emergente. Editora Novatec (2013)10. Martin Kleppmann. Designing Data-Intensive Applications: The Big Ideas Behind Reliable, Scalable, and Maintainable Systems. Editora O'Reilly (2017)11. Ralph Kimball, Margy Ross. The data warehouse toolkit: the definitive guide to dimensional modeling. John Wiley & Sons (2013)12. Ramez Elmasri, Shamkant B. Navathe. Sistemas de Banco de Dados. Editora Pearson (2019)13. Seiji Isotani, Ig Ibert Bittencourt. Dados Abertos Conectados. Editora Novatec (2015)
Sistematização da Prova Prática	