

PERÍODO: 2021.1		
CENTRO UNIVERSITÁRIO CTC DEPARTAMENTO DE INFORMÁTICA		
INF2031	Tópicos de Banco de Dados II Prof. Marco Antonio Casanova / Prof. Antonio Luz Furtado	
	CARGA HORÁRIA TOTAL: 45	CRÉDITOS: 3
	Título do Tópico: <i>Data Lakes</i> Pré-requisito:	
OBJETIVOS	Esta disciplina aborda tecnologias para construção de <i>Data Lakes</i> , armazenando grandes volumes de dados.	
EMENTA	Conceituação de <i>Data Lakes</i> ; tarefas típicas em um <i>Data Lake</i> ; importação de novos dados; inferência de tipos de dados; gerência de metadados; limpeza dos dados; integração dos dados; descoberta de dados; versionamento.	
PROGRAMA	<ul style="list-style-type: none"> • Introdução: Conceituação de <i>Data Lakes</i>; tarefas típicas em um <i>Data Lake</i> • Importação de novos dados: inclusão de dados estruturados no <i>Data Lake</i> e extração de dados estruturados de dados semi- ou não-estruturados e sua inclusão no <i>Data Lake</i> • Inferência de tipos de dados: mapeamento entre atributos de diferentes tabelas, anotação semântica e detecção de tipos de dados • Gerência de metadados: organização dos metadados importados para o <i>Data Lake</i> • Limpeza dos dados: correção de erros nos dados e preenchimento de dados ausentes, com o objetivo de melhorar a qualidade da análise dos dados • Integração dos dados: localização de conjuntos distintos de dados com semântica similar que possam ser unificados • Descoberta de dados: localização dos conjuntos de dados, armazenados no <i>Data Lake</i>, úteis a uma tarefa de análise 	

	<ul style="list-style-type: none">• Versionamento: manutenção no <i>Data Lake</i> das versões de uma mesma coleção de dados
AVALIAÇÃO	O desempenho do aluno será avaliado por seminários apresentados em sala e uma monografia preparada para o curso.
BIBLIOGRAFIA PRINCIPAL	Artigos selecionados
BIBLIOGRAFIA COMPLEMENTAR	Artigos selecionados