



## PROGRAMA DA DISCIPLINA

### DISCIPLINA:

INFRAESTRUTURA PARA GESTÃO DE DADOS

CODCRED	CARGA HORÁRIA	MÓDULO
98H00-04	60	60

### EMENTA:

Visão geral de sistemas de bancos de dados, suas principais funcionalidades e formas de uso. Entendimento da gerência de memória principal e secundária, de estruturas de acesso otimizado, e da sintonia de desempenho de bancos de dados. Estudo dos modelos de transações de bancos de dados, abordagens ACID e BASE, e de controle de concorrência e de recuperação após falhas. Compreensão das soluções para gestão de dados em Big Data. Visão geral sobre a infraestrutura existente para Big Data.

### OBJETIVOS:

O cumprimento da disciplina busca dar ao aluno, ao final do semestre, condições de:

1. Compreender os mecanismos básicos utilizados na implementação de SGBDs;
2. Compreender as diferentes formas de uso de bancos de dados, tanto convencionais como sem esquemas prévios;
3. Compreender os modelos de transação, de gestão de recursos, e de concorrência na implementação de SGBDs, e conhecer as principais técnicas de sintonia de desempenho;
4. Ter uma visão geral na gestão de dados não convencionais, multimídia e de redes sociais, e baseados em SGBDs sem esquemas prévios.

### CONTEÚDO:

#### Nº DA UNIDADE: 01

CONTEÚDO: Revisão de Sistemas de Gerência de Bancos de Dados (SGBD)

- 1.1 Noções gerais de bancos de dados
- 1.2 Modelos de dados
  - 1.2.1 Modelo de dados relacional
  - 1.2.2 Modelos de dados não relacionais
- 1.3 Paradigmas de processamento de dados
  - 1.3.1 Processamento transacional de dados
  - 1.3.2 Processamento analítico de dados





- 1.3.3 Processamento de big data
- 1.4 Arquitetura de SGBDs
- 1.5 Principais Serviços oferecidos por um SGBD
  - 1.5.1 Persistência de dados
  - 1.5.2 Acesso otimizado a dados
  - 1.5.3 Acesso disciplinado a dados

## **Nº DA UNIDADE: 02**

CONTEÚDO: Aspectos de implementação e desempenho

- 2.1 Interação com sistemas operacionais
- 2.2 Estruturas físicas de SGBDs
  - 2.2.1 Blocos arquivos de dados
  - 2.2.2 Gerência de acesso à memória
  - 2.2.3 Gerência de registros
  - 2.2.4 Estruturas de acesso otimizado.
    - 2.2.4.1 Árvores balanceadas (B-Trees)
    - 2.2.4.2 Acesso hash
    - 2.2.4.3 Outras estruturas
- 2.3 Sintonia de desempenho
  - 2.3.1 Aspectos de desempenho relacionados à infraestrutura
  - 2.3.2 Aspectos de desempenho relacionados à modelagem de dados
  - 2.3.3 Aspectos de sintonia relacionados ao acesso a dados
- 2.4 Sintonia de consultas SQL
  - 2.4.1 Mecanismo de resolução/otimização de consultas
  - 2.4.2 Otimização algébrica de consultas
  - 2.4.3 Algoritmos de acesso
  - 2.4.4 Análise e ajuste de planos de execução

## **Nº DA UNIDADE: 03**

CONTEÚDO: Paradigmas para gestão da consistência, disponibilidade e distribuição de dados

- 3.1 Consistência de dados em SGBDs
- 3.2 Abordagens ACID e BASE
  - 3.2.1 Teorema CAP
  - 3.2.2 Relaxando a consistência
  - 3.2.3 Relaxando a disponibilidade
- 3.3 Controle de concorrência
- 3.4 Recuperação após falhas
- 3.5 Modelos de distribuição de dados

## **Nº DA UNIDADE: 04**

CONTEÚDO: Infraestrutura para big data

- 4.1 Frameworks para processamento distribuído sobre clusters de computadores





- 4.1.1 Origens e evolução das soluções
- 4.1.2 Google File System e Hadoop Distributed File System
- 4.1.3 Comparação com arquiteturas clássicas para computação distribuída
- 4.2 Padrões e ferramentas de programação para processamento de big data
- 4.2.1 Arquitetura de implementação do padrão de programação Map-Reduce
- 4.2.2 Framework para computação distribuída Spark
- 4.3 SGBDs para soluções de big data
- 4.3.1 SGBDs chave-valor: Google BigTable, Hive
- 4.3.2 Outras abordagens para SGBDs: Cassandra, MongoDB, Neo4J
- 4.4 Tendências em infraestrutura para big data
- 4.4.1 Estruturas de dados para armazenamento e recuperação: Parquet, ORC, Solr
- 4.4.2 Outras tecnologias para infraestrutura de gestão de big data

#### **Nº DA UNIDADE: 05**

CONTEÚDO: Infraestrutura para suporte a gestão de dados

- 5.1 Ambientes de tecnologia da informação para gestão de dados
- 5.1.1 Plataforma on-premise e utility data center
- 5.1.2 Computação em nuvem
- 5.1.3 Soluções híbridas
- 5.2 Ambientes de computação na nuvem
- 5.2.1 Serviços oferecidos: infraestrutura, plataforma e SGBDs
- 5.2.2 Padrões de certificação para data centers e computação em nuvem
- 5.2.3 Modelos de contratação de infraestrutura para gestão de dados em nuvem
- 5.3 Bancos de dados em nuvem
- 5.3.1 Principais características e componentes
- 5.3.2 Estrutura e funcionamento de bancos de dados em nuvem
- 5.4 Critérios para seleção do modelo de infraestrutura para suporte a gestão de dados

#### **PROCEDIMENTOS E CRITÉRIOS DE AVALIAÇÃO:**

$$G1 = ( 0.35 * P + 0.65 * T )$$

Onde:

P – Prova versando sobre as unidades 01, 02 e 03.

T – Conjunto de trabalhos práticos abrangendo todas as unidades.

#### **BIBLIOGRAFIA BÁSICA:**

1. ELMASRI, R.; NAVATHE, S. B. Sistemas de banco de dados. 6ed., São Paulo: Pearson Brasil, 2012.
2. WHITE, T.. Hadoop: The definitive guide. O`Reilly Media, Inc., 2012.
3. SADALAGE, P. J.; FOWLER, M. NoSQL Essencial – Um Guia Conciso para o Mundo Emergente da Persistência Poliglota. São Paulo: Novatec, 2013.





**BIBLIOGRAFIA COMPLEMENTAR:**

1. GARCIA-MOLINA, H., ULLMAN, J.; WIDOM, J. Database systems: The Complete Book. 2 Ed. Prentice-Hall, 2014.
2. HUNG, P. (Editor) Big Data Applications and Use Cases. Cham(Suíça): Springer, 2016.
3. PIVERT, Olivier (Editor). NoSQL data model: trends and challenges. London: John Wiley, 2018.
4. RAMAKRISHNAN, R. Database management systems. 3rd Edition. Boston: McGraw-Hill, 2003.
5. SILBERSCHATZ, A.; KORTH, H. F.; SUDARSHAN, S. Sistema de bancos de dados. 6ed., Rio de Janeiro: Campus, 2012.

