

# AULA I - INTRODUÇÃO

PROFA. DRA. LEILA BERGAMASCO

CC6240 – Tópicos Avançados em Banco de Dados

# AGENDA

- Apresentação
- Métodos de Avaliação
- Introdução

# APRESENTAÇÃO

# QUEM SOU EU?



# QUEM SOU EU?

- Leila Cristina Carneiro Bergamasco

- Graduação em Sistemas de Informação (EACH-USP)

## Área de pesquisa

- Computação aplicada
  - Saúde
  - Sustentabilidade
- Processamento de imagens
- Inteligência Artificial
- Estatística

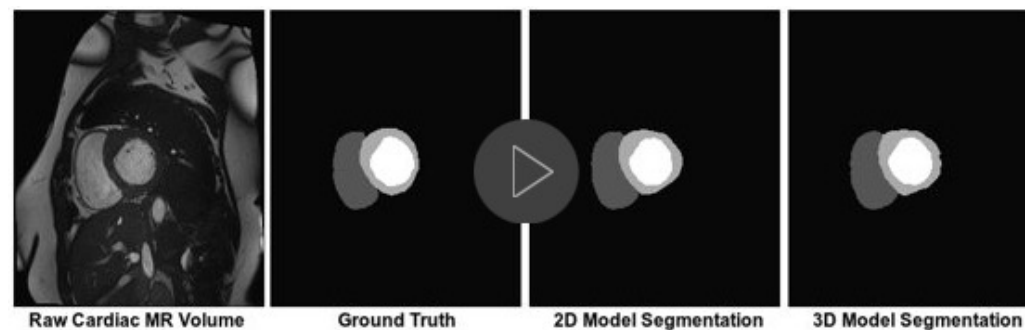
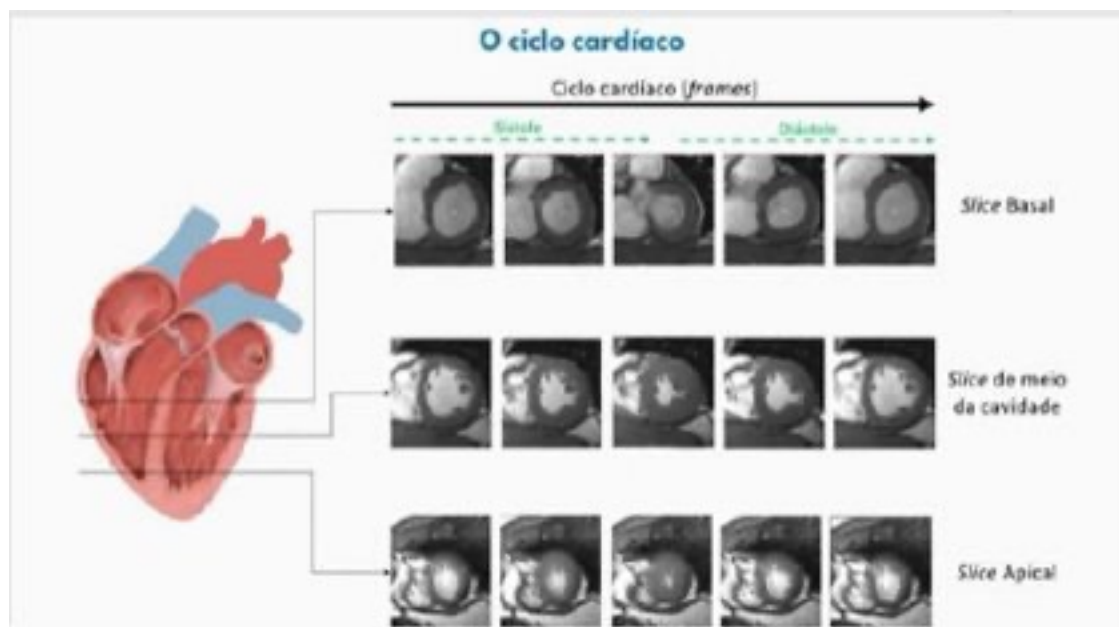
- Professora de Banco de Dados, Ciência de Dados, Algoritmos e afins na FEI



# ÁREAS DE PESQUISA



# ÁREAS DE PESQUISA



# ÁREAS DE PESQUISA





# ÁREAS DE PESQUISA



**MPF**  
**Ministério Público Federal**

# METODOLOGIA DE AVALIAÇÃO

# METODOLOGIA DA DISCIPLINA

- Plano da disciplina está no Moodle – tour pelo Moodle
  - Sujeito à alterações!
- Disciplina com aula expositiva e exercícios práticos
  - Material necessário:
    - PostgreSQL, caderno
  - Perguntem!
    - **Agendamentos** para resolução de dúvidas ou outros assuntos: [leila.cristina@fei.edu.br](mailto:leila.cristina@fei.edu.br)
    - **Não respondo mensagens no Moodle! Apenas por e-mail!**

# DINÂMICA DAS AULAS

- Aulas teóricas disponibilizadas em vídeo às quintas 22:40h
- Não é necessário vir presencialmente as aulas de sábado da disciplina CC6240
  - A presença de sábado será dada a todos os alunos.
- É necessário assistir e fazer anotações dos vídeos **ANTES** das aulas de Laboratório.
  - Serão tiradas dúvidas apenas!!
- Teremos atividades para entrega em todas as aulas!
  - Tempo será precioso para a realização e conclusão das atividades.
  - Feedbacks intermediários e/ou correções
  - Haverá tempo durante as aulas para vocês fazerem a maioria dos exercícios e o projeto, evitando “lições de casa”
    - Dessa forma, a presença de vocês é muito importante!

# METODOLOGIA DA DISCIPLINA

- Bibliografia

- ELMASRI, RAMEZ, SHAMKANT B. NAVATHE. Sistemas de banco de dados. Vol. 6. São Paulo: Pearson Addison Wesley, 2011.
- ABRAHAM SILBERSCHATZ, HENRY F. KORTH, S. SUDARSHAN. Sistema de Banco de Dados. 6. Campus



# METODOLOGIA DE AVALIAÇÃO

- Apresentações práticas
  - Desenvolvimento de uma oficina utilizando alguma tecnologia inovadora da área de banco de dados. Falaremos mais adiante.
- Avaliação Teórica
  - No fim do semestre abrangendo todo o conteúdo, com ênfase nas aulas teóricas
- Nota Final =  $(0.3*AP + 0.5*AT + 0.2*E)$ 
  - AP = Avaliação prática
  - AT = Avaliação teórica
  - E = Média de notas de exercícios

# OFICINA EM GRUPO

- Grupo de 3 - 4 alunos
- Oportunidade de aprendizado
- Melhorar portfólio/currículo
- Entregas pelo Moodle
- Plágio total ou parcial
  - Zero para todos os envolvidos
- Temas e enunciado completo serão disponibilizados nas próximas semanas.

# OFICINA EM GRUPO

- Realizar uma oficina de até 30 minutos sobre um tema inovador da área de banco de Dados
  - Banco de dados temporal, noSQL, streaming, video
  - Sugestões são bem-vindas
- Duas etapas
  - Teórica: pesquisa sobre o tema, aplicações – busca em bases científicas (20% da nota)
  - Prática: Elaborar uma oficina com os alunos de até 30 minutos. (80% da nota)
    - Se for fácil os alunos reproduzirem/acompanhar melhor
    - Demos, hands-on. O que vocês achem que ajuda a manter a turma engajada
- Notas
  - 60% da nota – Professora
  - 25% da nota – Avaliação dos outros grupos
  - 15% da nota – Qualidade das avaliação feita para os outros grupos

## OFICINA EM GRUPO



Nota da professora – 7

Média da nota dos outros grupos – 3

Média de qualidade das avaliações feitas (só deu 10, sem justificativa) – 1

Nota final =  $(0.6*7+0.25*3+0.15*1) = 5,1$

# OFICINA EM GRUPO

- Após as oficinas os grupos terão tempo para discutir e preencher o formulário de avaliação e submeter no Moodle.
  - Sem lição de casa, porém presença é fundamental!!
  - E se faltar? Sem penalização, porém o grupo é prejudicado por ter que fazer o formulário de avaliação com menos pessoas.
  - E se não vir ninguém? Sem nota para aquela apresentação → média da qualidade de avaliação cairá.
- Enunciado completo e temas nas próximas semanas!



# NOTA DE EXERCÍCIOS

- Toda a semana serão realizados exercícios de fixação, laboratório → entregue na aula.
- A média das notas comporá a nota de Exercícios (2 pt na média)

# AVALIAÇÃO TEÓRICA

- Moodle
- Todo o conteúdo do semestre

# OBJETIVO

- Preparar cidadãos e cidadãs conscientes sobre a complexidade dos desafios do mundo atual, no que tange principalmente à estruturação, armazenamento e recuperação dos dados.
- Desenvolver a habilidade de projetar soluções para problemas genéricos, analisando as especificidades dos dados a serem armazenados e as tecnologias disponíveis.
- Desenvolver a habilidade de diferenciar diferentes arquiteturas de armazenamento e indexação.
- Desenvolvendo
  - Foco
  - Estruturação de problemas
  - Abstração
  - Criatividade
  - Comunicação

# OBJETIVO

- Não é uma disciplina super fácil, mas não é a pior! 😊 Vamos passar juntos por alguns “gargalos”:
  - Cálculos
  - Abstração de problemas
  - Idealização de como fazer uma oficina
- É uma disciplina mais teórica que prática!
  - 70% teórica
  - 30% prática.

# PALESTRANTES EXTERNOS!

- Em geral 1 ou 2 palestras de pessoas de mercado que trabalham com banco de dados



# O QUE VAMOS PRECISAR PARA ESSA DISCIPLINA?

- Banco de dados relacional
  - Alguns conceitos vocês poderão utilizar um BD relacional para testar/praticar
  - Nas aulas vou utilizar o PostgreSQL, mas podem escolher os que vocês tem mais familiaridade.
- Paciência
  - A parte de otimização é bem pesada e tem alguns cálculos.
- Criatividade
  - Para as oficinas!

# TÓPICOS AVANÇADOS EM BANCO DE DADOS



# A DAY IN DATA

The exponential growth of data is undisputed, but the numbers behind this explosion – fuelled by internet of things and the use of connected devices – are hard to comprehend, particularly when looked at in the context of one day

**500m**

tweets are sent every day

Twitter



**4PB**

of data created by Facebook, including

**350m** photos

**100m** hours of video watch time

Facebook Research

**294bn**

billion emails are sent

Radicati Group

**320bn**

emails to be sent each day by 2021

**306bn**

emails to be sent each day by 2020

**3.9bn**

people use emails

**4TB**

of data produced by a connected car

Intel

## DEMYSTIFYING DATA UNITS

From the more familiar 'bit' or 'megabyte', larger units of measurement are more frequently being used to explain the masses of data

Unit	Value	Size
<b>b</b> bit	0 or 1	1/8 of a byte
<b>B</b> byte	8 bits	1 byte
<b>KB</b> kilobyte	1,000 bytes	1,000 bytes
<b>MB</b> megabyte	1,000 <sup>2</sup> bytes	1,000,000 bytes
<b>GB</b> gigabyte	1,000 <sup>3</sup> bytes	1,000,000,000 bytes
<b>TB</b> terabyte	1,000 <sup>4</sup> bytes	1,000,000,000,000 bytes
<b>PB</b> petabyte	1,000 <sup>5</sup> bytes	1,000,000,000,000,000 bytes
<b>EB</b> exabyte	1,000 <sup>6</sup> bytes	1,000,000,000,000,000,000 bytes
<b>ZB</b> zettabyte	1,000 <sup>7</sup> bytes	1,000,000,000,000,000,000,000 bytes
<b>YB</b> yottabyte	1,000 <sup>8</sup> bytes	1,000,000,000,000,000,000,000,000 bytes

\*A lowercase "b" is used as an abbreviation for bits, while an uppercase "B" represents bytes.

**65bn**

messages sent over WhatsApp and two billion minutes of voice and video calls made

Facebook

Searches made a day

**5bn**

Searches made a day from Google

**3.5bn**

Smart Insights

**463EB**

of data will be created every day by 2025

IDC

**95m**

photos and videos are shared on Instagram

Instagram Business

**28PB**

to be generated from wearable devices by 2020

Statista

## ACCUMULATED DIGITAL UNIVERSE OF DATA

**4.4ZB**

**44ZB**

PwC

2019

2020

# POR QUE APRENDER TÓPICOS AVANÇADOS EM BD?

- Dados heterogêneos



Sensores móveis



FACEBOOK CRESCE  
COM A  
PUBLICAÇÃO DE  
250 MILHÕES  
DE FOTOS/DIA

Redes sociais



Vídeos de  
vigilância

Renderização de Vídeos



LER MEDIDORES  
A CADA 15 MIN  
GERA 3,000X  
MAIS DADOS



Smart Grids



Exploração  
Geofísica

Imagens médicas



Sequenciamento de gens



O CUSTO DE SEQUENCIAR  
UM GENOMA CAIU  
DE \$100M EM 2001  
PARA \$10K EM 2011

# PROGRAMA RESUMIDO

- Características físicas de um Banco de Dados
- Indexação, hashing e otimização
- Diferentes tecnologias de Banco de Dados
- Data mining\*