

# Bancos de Dados no BCC (e em todos os lugares!)

Kelly Rosa Braghetto

DCC-IME-USP

12 de maio de 2016

# O que é um **banco de dados**?

**Você sabe citar exemplos?**

# O que é um **banco de dados**?

## Exemplos

- Uma agenda de telefones e endereços de seus contatos
- O catálogo com as informações do acervo de uma biblioteca
- Os dados de imposto de renda da Receita Federal
- Os registros de matrículas e notas dos alunos de uma universidade
- As informações sobre o estoque e as vendas de uma loja
- Os prontuários dos pacientes de um hospital
- Os registros meteorológicos coletados na cidade de São Paulo
- ...

# O que é um **banco de dados**?

## Definições:

- **Banco de dados** – coleção de dados relacionados
- **Dados** – fatos conhecidos que podem ser registrados e que possuem significado implícito

**Referência:** “Sistemas de Bancos de Dados” (6 a edição), Elmasri e Navathe. Pearson, 2010.

# O que é um **banco de dados**?

## Definições:

- **Banco de dados** – coleção de dados relacionados
- **Dados** – fatos conhecidos que podem ser registrados e que possuem significado implícito

**Definições genéricas demais!**

**Referência:** “Sistemas de Bancos de Dados” (6 a edição), Elmasri e Navathe. Pearson, 2010.

# Propriedades implícitas de um banco de dados (BD)

- 1 Representar um aspecto do mundo real
- 2 Ser uma **coleção lógica e coerente de dados** com algum **significado inerente**  
Uma coleção “aleatória” de dados não é um BD!
- 3 Ser projetado, construído e povoado com dados que possuem um objetivo específico  
Um BD deve possuir um grupo de usuários em potencial e algumas aplicações pré-concebidas, nas quais esses usuários estão interessados

# Disciplinas do DCC-IME-USP sobre Bancos de Dados

- **MAC0350 Introdução ao Desenvolvimento de Sistemas de Software** [Obrigatória]
- **MAC0426 Sistemas de Bancos de Dados** [Eletiva]
- **MAC0439 Laboratório de Banco de Dados** [Eletiva]
- **MAC0459 Ciência e Engenharia de Dados** [Eletiva]

# Professores de Bancos de Dados do DCC-IME-USP

- João Eduardo Ferreira – [jef@ime.usp.br](mailto:jef@ime.usp.br)
- Kelly Rosa Braghetto – [kellyrb@ime.usp.br](mailto:kellyrb@ime.usp.br)
- Marcelo Finger – [mfinger@ime.usp.br](mailto:mfinger@ime.usp.br)





# Softwares para a manutenção de bancos de dados

Um dado BD informatizado pode ser criado e mantido por:

- um programa de aplicação desenvolvido especificamente para essa tarefa

**ou**

- um **Sistema de Gerenciamento de Banco de Dados (SGBD)**

Sistema de software de propósito geral que facilita o processo de **definição, construção, manipulação e compartilhamento** de BDs entre vários usuários e aplicações

# Um sistema de banco de dados monousuário



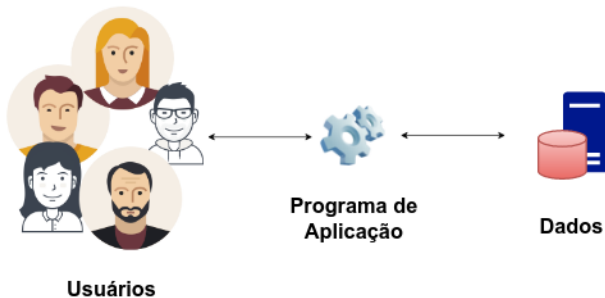
# Um sistema de banco de dados monousuário



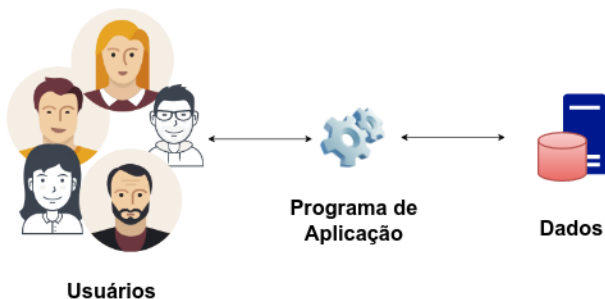
## Dificuldades:

- Manter a consistência dos dados
- Possibilitar a recuperação e manipulação eficiente de dados
- Proteger os dados de acessos indevidos
- Recuperar-se de falhas

# Um sistema de banco de dados multiusuário



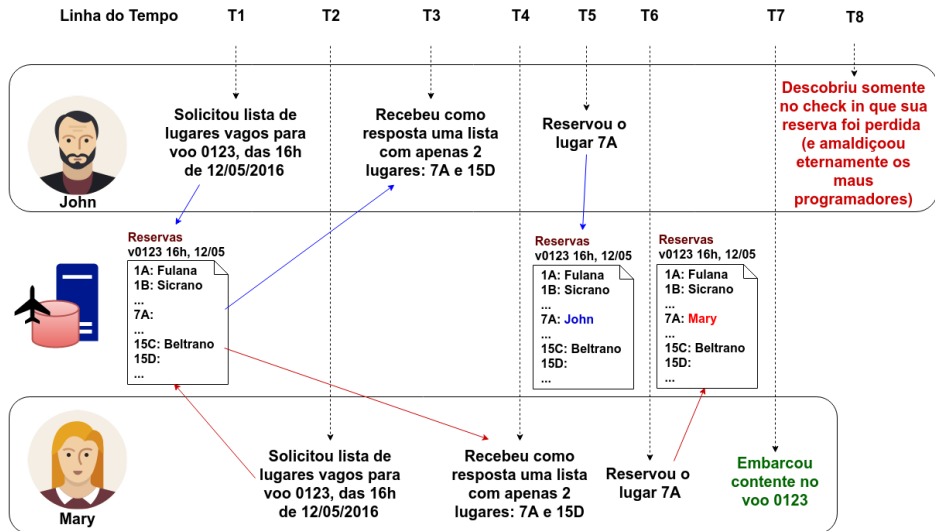
# Um sistema de banco de dados multiusuário



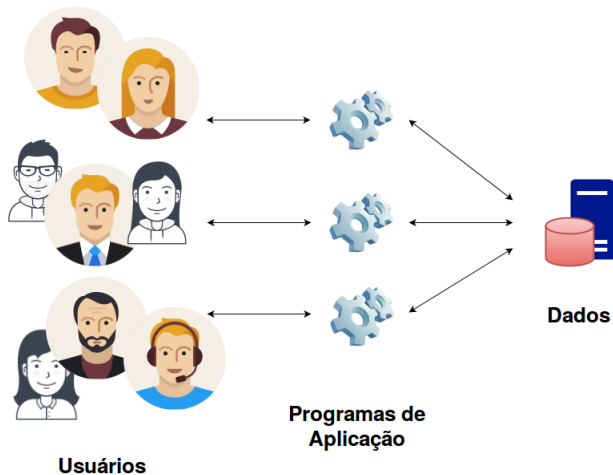
## Dificuldades adicionais:

- Gerenciar as necessidades/permisões dos diferentes tipos de usuários
- Evitar conflitos causados por acessos concorrentes aos dados

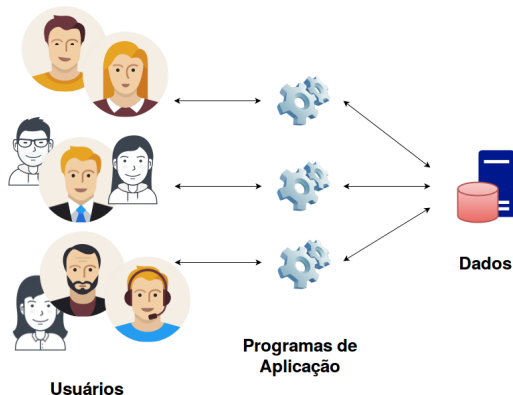
# O problema do acesso concorrente aos dados



# Outro sistema de banco de dados multiusuário



# Outro sistema de banco de dados multiusuário

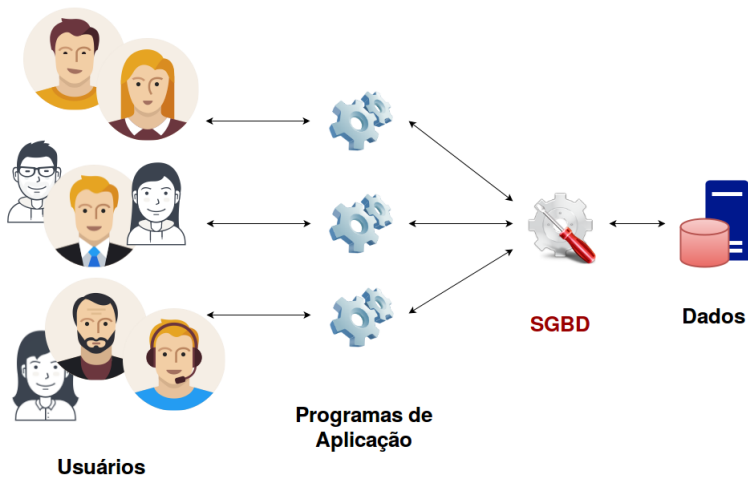


## Dificuldades adicionais:

- Manter detalhes sobre a organização dos dados no código de várias aplicações ao mesmo tempo



# Um sistema de banco de dados com SGBD



# Funções Importantes de um SGBD

## (Sistema Gerenciador de Banco de Dados)

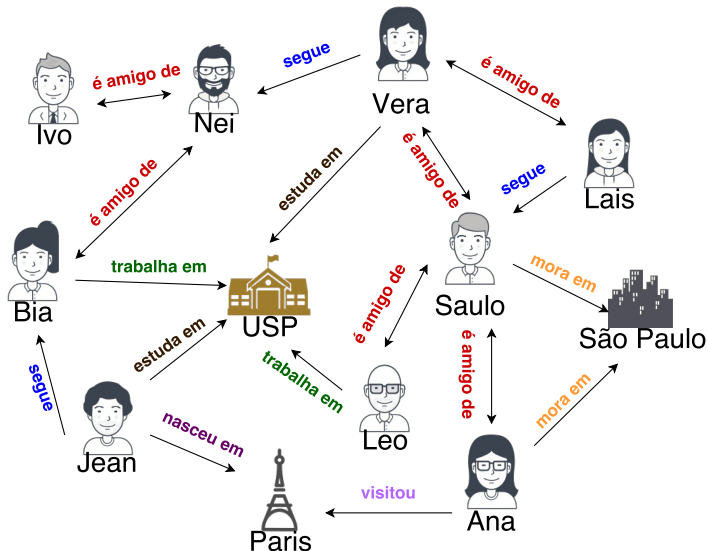
- Manter os dados por um longo período de tempo
- Garantir a consistência dos dados
- Prover eficiência no acesso aos dados
- Proteger as dados contra acessos não autorizados
- Proteger os dados contra falhas de hardware ou software
- Prover isolamento entre programas e dados
- Prover suporte a visões múltiplas dos dados
- Possibilitar o compartilhamento de dados e processar transações multiusuários

# Funções Importantes de um SGBD

## (Sistema Gerenciador de Banco de Dados)

- Manter os dados por um longo período de tempo
- Garantir a consistência dos dados
- Prover eficiência no acesso aos dados
- Proteger as dados contra acessos não autorizados
- Proteger os dados contra falhas de hardware ou software
- **Prover isolamento entre programas e dados**
- Prover suporte a visões múltiplas dos dados
- Possibilitar o compartilhamento de dados e processar transações multiusuários

## Um exemplo motivacional – dados de uma “rede social”



# Um exemplo motivacional – dados de uma “rede social”

Queremos responder perguntas como:

- “Quem são os amigos da Laís?”
  - “Quem são os amigos indiretos da Laís?”
  - “Existe alguém que não tem nenhum amigo?”
  - “Quantos amigos uma pessoa tem em média?”
- 
- De que formas podemos armazenar os dados do grafo em um arquivo em disco?
  - Essas formas facilitam a realização das consultas descritas acima?

# Representação matricial

	Ana	Bia	Ivo	Jean	Laís	Leo	Nei	Saulo	Vera
Ana								1	
Bia							1		
Ivo							1		
Jean		2							
Laís								2	1
Leo								1	
Nei		1	1						
Saulo	1					1			1
Vera					1		2	1	

**1: é amigo de**

**2: segue**

# Representação matricial

	Ana	Bia	Ivo	Jean	Laís	Leo	Nei	Saulo	Vera
Ana								1	
Bia							1		
Ivo							1		
Jean		2							
Laís								2	1
Leo								1	
Nei		1	1						
Saulo	1					1			1
Vera					1		2	1	

**1: é amigo de**

**2: segue**

## Problemas:

- Não é viável para uma rede social com muitas pessoas (desperdiça espaço em disco)
- Se cada linha da matriz corresponde a um registro (ou linha) do arquivo, a inserção de uma nova pessoa na rede social demandaria a reorganização do arquivo todo

# Representação com matriz esparsa

<b>Pessoa</b>	<b>Lista de Amigos</b>
Ana	<b>Saulo</b>
Bia	<b>Nei</b>
Ivo	<b>Nei</b>
Jean	
Laís	<b>Vera</b>
Leo	<b>Saulo</b>
Nei	<b>Bia, Ivo</b>
Saulo	<b>Ana, Leo, Vera</b>
Vera	<b>Laís, Saulo</b>



# Representação com matriz esparsa

<b>Pessoa</b>	<b>Lista de Amigos</b>
Ana	Saulo
Bia	Nei
Ivo	Nei
Jean	
Láis	Vera
Leo	Saulo
Nei	Bia, Ivo
Saulo	Ana, Leo, Vera
Vera	Láis, Saulo

## Problemas:

- Se cada linha da matriz corresponde a um registro (ou linha) do arquivo, a inserção de um novo amigo para uma pessoa já existente na rede social pode demandar a reorganização do arquivo

# Representação tabular

<b>Pessoa</b>	<b>Amigo 1</b>	<b>Amigo 2</b>	<b>Amigo 3</b>	<b>Amigo 4</b>	<b>Amigo 5</b>
<b>Ana</b>	<b>Saulo</b>				
<b>Bia</b>	<b>Nei</b>				
<b>Ivo</b>	<b>Nei</b>				
<b>Jean</b>					
<b>Láís</b>	<b>Vera</b>				
<b>Leo</b>	<b>Saulo</b>				
<b>Nei</b>	<b>Bia</b>	<b>Ivo</b>			
<b>Saulo</b>	<b>Ana</b>	<b>Leo</b>	<b>Vera</b>		
<b>Vera</b>	<b>Láís</b>	<b>Saulo</b>			

# Representação tabular

Pessoa	Amigo 1	Amigo 2	Amigo 3	Amigo 4	Amigo 5
Ana	Saulo				
Bia	Nei				
Ivo	Nei				
Jean					
Láís	Vera				
Leo	Saulo				
Nei	Bia	Ivo			
Saulo	Ana	Leo	Vera		
Vera	Láís	Saulo			

## Problemas:

- A quantidade de amigos por pessoa é limitada

# Outra representação tabular

<b>Pessoa</b>	<b>Amigo</b>
Ana	Saulo
Bia	Nei
Ivo	Nei
Laís	Vera
Leo	Saulo
Nei	Bia
Nei	Ivo
Saulo	Ana
Saulo	Leo
Saulo	Vera
Vera	Laís
Vera	Saulo

## Outra representação tabular

<b>Pessoa</b>	<b>Amigo</b>
Ana	Saulo
Bia	Nei
Ivo	Nei
Láis	Vera
Leo	Saulo
Nei	Bia
Nei	Ivo
Saulo	Ana
Saulo	Leo
Saulo	Vera
Vera	Láis
Vera	Saulo

### Problemas:

- A busca por informações sobre os amigos (diretos e indiretos) de uma pessoa fica mais demorada

# Abstração de dados

Oferecer abstração de dados é uma característica fundamental dos bancos de dados, ocultando de usuários e aplicações detalhes sobre a organização e armazenamento dos dados

A abstração é feita por meio de modelos de dados:

- **Modelo de dados** – conjunto de conceitos usados para descrever a *estrutura* de um banco de dados + *operações* básicas para a recuperação e atualização de dados do banco
- **Estrutura de um banco de dados** – define os tipos de dados, relacionamentos e restrições que se aplicam aos dados

## Alguns exemplos de modelos bastante usados (e seus SGBDs)

- **Modelo Relacional** (PostgreSQL, MySQL, Oracle, MS SQL Server, DB2, Sybase)
- **Modelo Orientado a Objetos** (db4o, Caché)
- **Modelo Orientado a Grafos** (Neo4J, Titan)
- **Modelo Chave-Valor** (Redis, Ryak)
- **Modelo Orientado a Documentos** (MongoDB, CouchDB )
- **Modelo Orientado a Colunas** (MonetDB, Cassandra)

# Exemplo de BD no modelo relacional: alunos e disciplinas

## ALUNO

Nome	Numero_aluno	Tipo_aluno	Curso
Silva	17	1	CC
Braga	8	2	CC

## DISCIPLINA

Nome_disciplina	Numero_disciplina	Creditos	Departamento
Introd. à ciência da computação	CC1310	4	CC
Estruturas de dados	CC3320	4	CC
Matemática discreta	MAT2410	3	MAT
Banco de dados	CC3380	3	CC

## TURMA

Identificacao_turma	Numero_disciplina	Semestre	Ano	Professor
85	MAT2410	Segundo	07	Kieber
92	CC1310	Segundo	07	Anderson
102	CC3320	Primeiro	08	Carlos
112	MAT2410	Segundo	08	Chang
119	CC1310	Segundo	08	Anderson
135	CC3380	Segundo	08	Santos

## HISTORICO\_ESCOLAR

Numero_aluno	Identificacao_turma	Nota
17	112	B
17	119	C
8	85	A
8	92	A
8	102	B
8	135	A

## PRE\_REQUISITO

Numero_disciplina	Numero_pre_requisito
CC3380	CC3320
CC3380	MAT2410
CC3320	CC1310



# Exemplo de metadados mantidos por um SGBD

## RELAÇOES

Nome_relacao	Numero_de_colunas
ALUNO	4
DISCIPLINA	4
TURMA	5
HISTORICO_ESCOLAR	3
PRE_REQUISITO	2

## COLUNAS

Nome_coluna	Tipo_de_dado	Pertence_a_relacao
Nome	Caractere (30)	ALUNO
Numero_aluno	Caractere (4)	ALUNO
Tipo_aluno	Inteiro (1)	ALUNO
Curso	Tipo_curso	ALUNO
Nome_disciplina	Caractere (10)	DISCIPLINA
Numero_disciplina	XXXXNNNN	DISCIPLINA
....	....	....
....	....	....
....	....	....
Numero_pre_requisito	XXXXNNNN	PRE-REQUISITO

## Representação de um registro de aluno

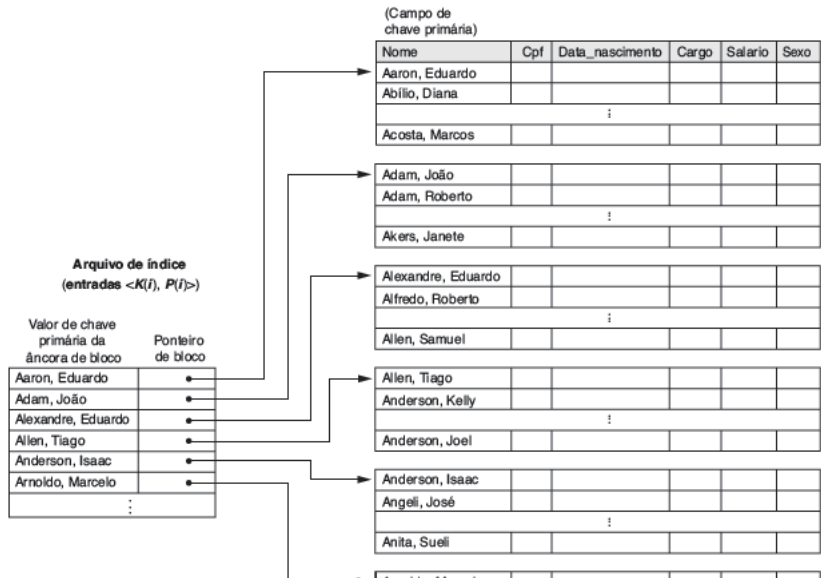
Nome do item de dados	Posicionamento inicial no registro	Tamanho em caracteres (bytes)
Nome	1	30
Numero_aluno	31	4
Tipo_aluno	35	1
Curso	36	4

# Linguagens de definição de esquemas e manipulação e consulta de dados

## Exemplo de comandos em SQL (*Structured Query Language*)

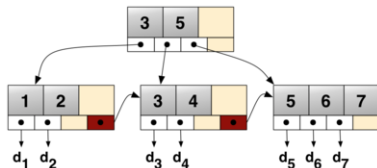
```
CREATE TABLE Disciplina (  
Nome_disciplina    CHAR(30),  
Numero_disciplina  CHAR(7),  
Creditos           INT,  
Departamento      CHAR(3)  
);  
  
INSERT INTO Disciplina  
VALUES ('Sistemas de Bancos de Dados', 'MAC0426', 4, 'DCC');  
  
SELECT Nome_disciplina, Creditos  
FROM   Disciplina  
WHERE  Departamento = 'MAT' AND Creditos = 6;
```

# Uso de índices para acelerar as consultas

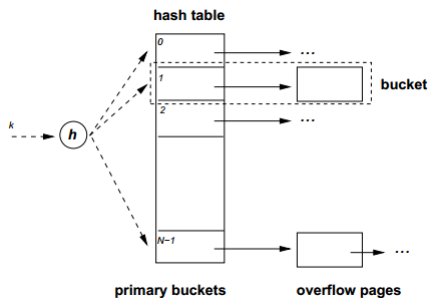


# Uso de índices para acelerar as consultas

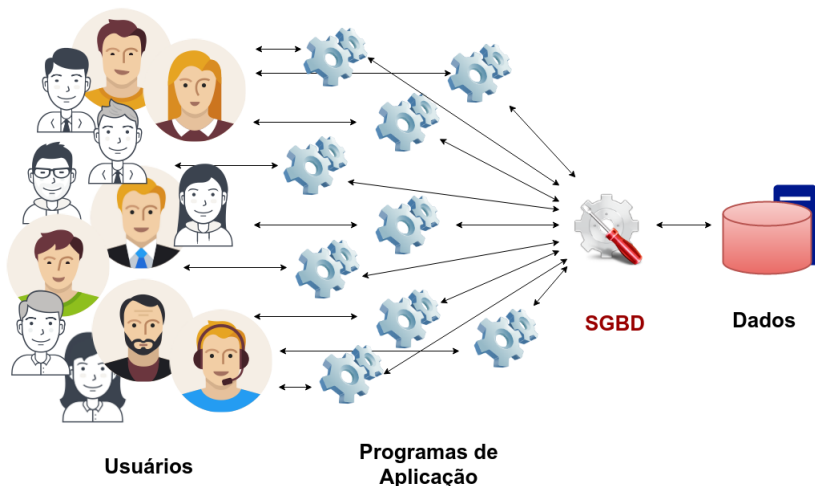
## Árvores B+



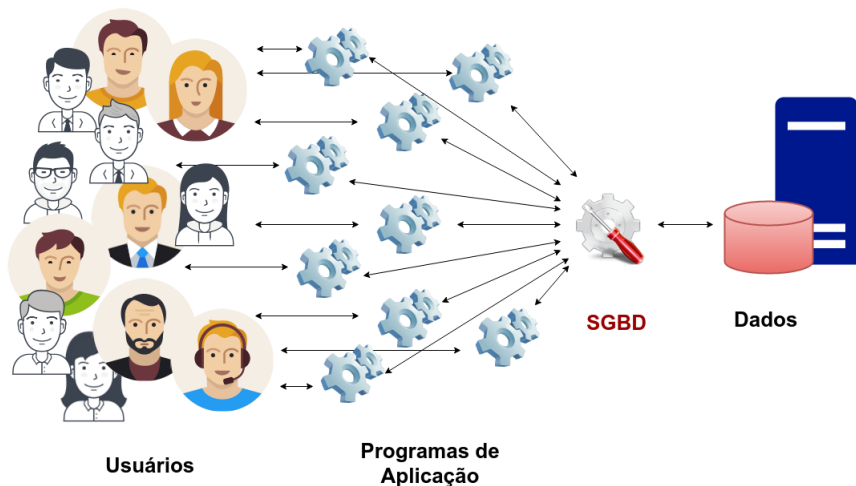
## Tabelas Hash



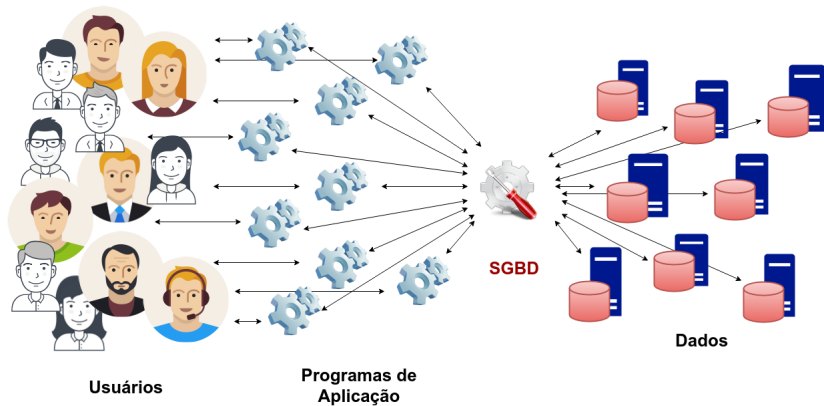
# O que fazer quando os dados ou os acessos crescem demais?



# SGBD centralizado, escalabilidade vertical



# SGBD distribuído, escalabilidade horizontal



# MAC0350 Introdução ao Desenvolvimento de Sistemas de Software

## Programa Resumido

- Técnicas p/ coleta de requisitos, modelagem, projeto e implementação de sw
- Conceitos de projetos de sistemas avançados.
- Qualidade , gerência e manutenção de software
- Usabilidade e experiência do usuário
- Arquitetura de software, projeto orientado a objetos, padrões de projeto
- Verificação e validação com testes automatizados
- **Persistência de dados: projeto conceitual, lógico e físico de BDs**
- **Modelo entidade relacionamento e modelo relacional: álgebra relacional, cálculo relacional e linguagem SQL**

**Pré-requisito:** MAC0316 Conceitos Fundamentais de Linguagens de Programação



# MAC0426 Sistemas de Bancos de Dados

## Nova Ementa!

### Programa Resumido

- Fundamentos da arquitetura dos sistemas SGBDs relacionais
  - armazenamento, indexação, processamento e otimização de consultas, controle de concorrência, recuperação de falhas
- Dados semiestruturados e suas linguagens de consulta
- Introdução aos bancos de dados paralelos e distribuídos
- Introdução a bancos de dados não relacionais
  - chave-valor, orientados a documentos, orientados a colunas e orientados a grafos
- Paradigma de programação *map reduce*

**Pré-requisito:** MAC0350 Introdução ao Desenvolvimento de Sistemas de Software

# MAC0439 Laboratório de Banco de Dados

## Programa Resumido

- Especificação e implementação de um BD com seus procedimentos de inserção, eliminação, atualização e consulta
- Ferramentas de apoio para o desenvolvimento do projeto lógico e físico de BDs
- Gerenciamento de BDs relacionais por meio de comandos da linguagem SQL
  - definição de dados e restrições; modificações e consultas sobre os dados
  - visões, gatilhos e procedimentos armazenados
  - segurança; transações; índices
- Sintonia fina de BDs relacionais
- Integração de BDs a sistemas de software

**Pré-requisito:** MAC0350 Introdução ao Desenvolvimento de Sistemas de Software

# MAC0459 Ciência e Engenharia de Dados

## Nova Ementa!

### Programa Resumido

- Processo de descoberta do conhecimento (KDD – *knowledge-discovery in databases*)
- Tratamento, representação e qualificação de grande volumes de dados
- Armazém de dados e modelos multidimensionais
- Noções de Redes de Petri e Álgebra de Processos
- Modelagem e implementação de workflows científicos e processos de negócio
- Indexação e recuperação de grande volumes de dados
- Análise exploratória de dados
- Redes complexas e bancos de dados orientado a grafos

**Pré-requisito:** MAC0350 Introdução ao Desenvolvimento de Sistemas de Software

# Trilha de Ciências de Dados

## Ser aprovado em todas

- MAE0221 Probabilidade I
- MAE0312 Introdução aos Processos Estocásticos
- MAC0460 Aprendizagem Computacional: Modelos, Algoritmos e Aplicações,
- MAC0317 Algoritmos para Processamento de Áudio, Imagem e Vídeo,
- **MAC0426: Sistemas de Bancos de Dados**
- MAC0431 Introdução à Computação Paralela e Distribuída

## Ser aprovado em 1

- MAC0315 Programação Linear
- MAC0325 Otimização Combinatória
- MAC0427 Programação não-Linear

## TCC em Ciências de Dados

- O aluno elege um orientador de TCC na área de Ciências de Dados
- Os dois elegem um tema para o TCC e pelo menos duas disciplinas relacionadas com o tema escolhido, que devem ser cursadas pelo aluno
- Possíveis temas incluem áreas de aplicação (e.g., Bioinformática, Economia, etc.) e de fundamentos (e.g., aprendizado computacional, processamento de sinais e imagens, etc.)

# Trilha de Sistemas de Software

## Bancos de Dados

Ser aprovado em 2

- **MAC0439** Sistemas de Bancos de Dados
- **MAC0426** Laboratório de Banco de Dados
- **MAC0459** Ciência e Engenharia de Dados

## Desenvolvimento de Software

Ser aprovado em todas

- **MAC0332** Engenharia de Software
- **MACxxxx** Laboratório de Métodos Ágeis
- **MACxxxx** Técnicas de Programação 2
- **MAC0413** Tópicos Avançados de Programação OO

## Sistemas Paralelos e Distribuídos

Ser aprovado em 3

- **MACxxxx** Arquitetura de Computadores
- **MAC0463** Computação Móvel
- **MACxxxx** Construção de Software como Serviço em Computação em Nuvem
- **MACxxxx** Desenvolvimento para Web
- **MACxxxx** Programação Concorrente e Paralela
- **MACxxxx** Redes de Computadores e Sistemas Distribuídos

Cursando as destacadas, você terá: uma formação em BD bacana, um certificado da Trilha de Sistemas e “espaço” para outras 4 eletivas em outros tópicos de interesse.

# Para entender melhor a implementação de índices e a análise de seus desempenhos...

- **MAC0121 Algoritmos e Estruturas de Dados I**  
Obrigatória, 2o. semestre, Pré-requisito: MAC0110  
(...) Árvores binárias. (...) Algoritmos de ordenação. Processamento elementar de texto. Tabelas de símbolos elementares, incluindo árvores binárias de busca.
- **MAC0323 Algoritmos e Estruturas de Dados II**  
Obrigatória, 3o. semestre, Pré-requisito: MAC0121, MAC0216  
(...) Tabelas de símbolos: árvores de busca balanceadas, tabelas de espalhamento (hashing). (...)
- **MAC0338 Análise de Algoritmos**  
Obrigatória, 4o. semestre, Pré-requisito: MAC0121  
Análise da correção e do desempenho de algoritmos. (...)

# Para entender melhor a base do modelo relacional e as linguagens de consulta de dados...

- **MAC0329 Álgebra Booleana e Circuitos Digitais**

Obrigatória, 1o. semestre

(...) Álgebra booleana: definição axiomática, exemplos (álgebra de conjuntos, cálculo proposicional, funções lógicas), propriedades, e ordens parciais em álgebras booleanas.

- **MAC0239 Introdução à Lógica e Verificação de Programas**

Obrigatória, 2o. semestre

Lógica Formal: cálculo proposicional, sintaxe, semântica, métodos de prova; cálculo de predicados de primeira ordem. (...)

# Para entender melhor projeto e desenvolvimento de BDs...

- **MACxxxx Engenharia de Software**  
Eletiva, Trilha de Sistemas, Pré-requisito: MAC0350
- **MACxxxx Técnicas de Programação 2**  
Eletiva, Trilha de Sistemas, Pré-requisito: MAC0216



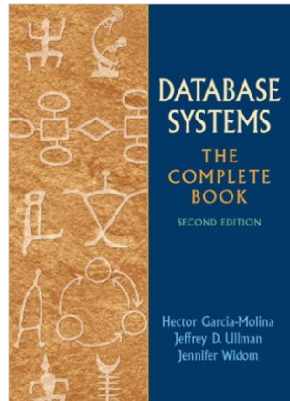
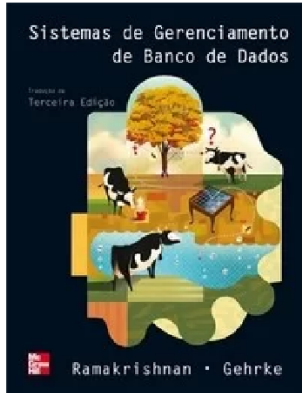
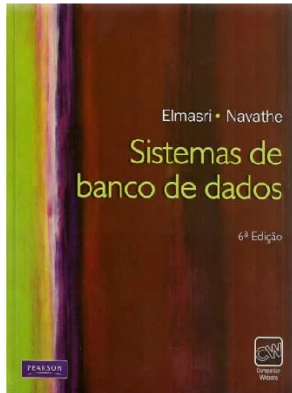
# Para entender melhor controle de concorrência e arquiteturas distribuídas...

- **MACxxxx Programação Concorrente e Paralela**  
Eletiva, Trilha de Sistemas, Pré-requisitos: MAC0121, MAC0216
- **MACxxxx Redes de Computadores e Sistemas Distribuídos**  
Eletiva, Trilha de Sistemas, Pré-requisito: MAC0323

## Saindo das obrigatórias e da trilha de sistemas...

- **MAC0333 Armazenamento e Recuperação de Informação**  
Eletiva, Pré-requisito: MAC0323
- **MAC0444 Sistemas Baseados em Conhecimento**  
Eletiva, Pré-requisito: MAC0323, MAC0239

# Bibliografia Básica Recomendada



# Bibliografia Básica Recomendada

