

Motivação e Conceitos Fundamentais

Vanessa Braganholo

vanessa@ic.uff.br

Conceitos Básicos

Afinal, o que é Banco de Dados?

Bom Dia Cidade – São Carlos/Araraquara >

Covid-19: Sumaré registra 741 novos casos após atualização no banco de dados do Estado

1 min Exibição em 20 ago 2020

Ataque arrasa bancos de dados e deixa uma mensagem: “miau!”

NOTÍCIAS

Dilma sanciona lei que cria banco de dados genéticos de criminosos

Política

Lava-jato fluminense diz que acesso da PGR aos bancos de dados da força-tarefa seria uma 'devassa'

ADVOCACIA DIGITAL

Harvard decide criar banco de dados de opiniões das cortes dos EUA

A informação foi divulgada pelo diretor do Laboratório de Inovação da Biblioteca da Faculdade

Bom Dia GO >

SSP lança banco de dados digitalizado para ajudar na busca por pessoas desaparecidas

3 min Exibição em 18 ago 2020

No mundo científico...

NCBI Influenza Virus Resource
Information, Search and Analysis

Home Search Site Map Flu home Database Alignment Genome Set Tree BLAST FTP Help Contact us

Main Page>>Database

What are you looking for? Select one name each from the lists provided, and/or fill in the boxes. Multiple queries can be built by clicking the "Add to Query Builder" button every time a new query is made, and queries in any combination from the Query Builder can be selected to get sequences in the database. An advanced search tool is available [here](#).

show ☐ Protein sequence ☒ Coding region ☐ Nucleotide sequence

Virus Species Host Country/Region Segment

any any 1 (PB2)

Influenzavirus A Avian Africa 2 (PB1)

Influenzavirus B Camel Asia 3 (PA)

Influenzavirus C Canine Europe 4 (HA)

Subtype From year To year Min. length Max. length

H5N1 2005 2005 Search by a string [Help](#)

☒ Pull-length sequences only [Help](#) ☐ Sequences from the FLU project only [Help](#) ☐ Include Lab strains [Help](#)

Add to Query Builder Clear selection Find sequence by Accession

Query Builder

Get sequences

☒	Virus species	Host	Country/Region	Segment	Subtype	Year	Length	Key word	Pull-length	NSAID project	Include Lab strains	Number of sequences
☒	Influenzavirus A	Avian	Asia	4 (HA)	H5N1	2005-2005	-		yes	no	no	85

Get sequences



Illustration by Sandbox Studio, Chicago with Ana Kova

An astronomical data challenge

02/19/19 | By Manuel Gnida

The Large Synoptic Survey Telescope will manage unprecedented volumes of data produced each night.



**Telescope
Data Center**
SMITHSONIAN ASTROPHYSICAL OBSERVATORY

Astronomical Data Sources on the Web

Para entender o que é um banco de dados, vamos começar com um exemplo simples...

Problema...

“Eu preciso calcular a média das notas de uma turma de graduação que eu estou ministrando.”

Opa... eu manjo de Python!

```
t = input("Entre com o número de alunos da turma: ")
notas = []
//armazena notas de 3 provas dos alunos de uma turma em uma matriz
for i in range(t):
    notas_aluno = []
    notas_aluno.append(input("Entre com a nota da P1 do aluno " + i + ":"))
    notas_aluno.append(input("Entre com a nota da P2 do aluno " + i + ":"))
    notas_aluno.append(input("Entre com a nota da P3 do aluno " + i + ":"))
//calcula a média e armazena na quarta coluna da matriz
    notas_aluno.append((notas[i][0] + notas[i][1] + notas[i][2])/3)
    notas.append(notas_aluno)
...
```



O programinha funciona, e agora?

Mas, na hora de lançar as notas no idUFF....

- me dou conta de que vou ter que digitar todas as notas novamente!

Por que?

- Porque o programa não “salva” os dados digitados pelo professor

A cada execução, a memória está zerada...

- Nada é persistido!

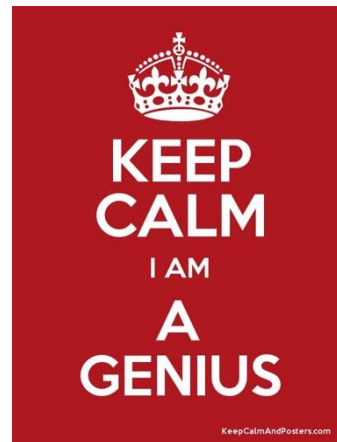


Solução?

Vamos gravar as notas em arquivos!

Digito uma vez só e posso ler múltiplas vezes

As notas ficam associadas à matrícula do aluno e ao Ano/Semestre que ele cursou a disciplina!



Arquivo de Notas

Matrícula	Ano	Semestre	P1	P2	P3
1001001	2019	1	10,0	8,0	7,0
1002001	2019	1	7,0	5,5	1,2
1003001	2019	1	4,0	7,0	8,0
1001002	2019	2	8,0	4,5	7,1
1002002	2019	2	7,5	6,5	7,0
1001003	2020	1	8,4	5,6	4,0
1002003	2020	1	7,0	4,5	9,0

Mas...

Depois de 20 turmas o arquivo fica grande...

Difícil de manipular

Será que podemos ter problemas?

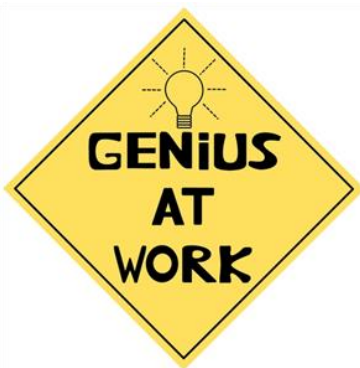
Cenário #1

Um aluno se candidatou para ser bolsista de IC

Quero consultar a nota dele para decidir pela aceitação

O que tenho que fazer?

- Escrever um programa para ler o arquivo sequencialmente!



Matrícula a procurar: 1001003



Matrícula	Ano	Semestre	P1	P2	P3
1001001	2019	1	10,0	8,0	7,0
1002001	2019	1	7,0	5,5	1,2
1003001	2019	1	4,0	7,0	8,0
1001002	2019	2	8,0	4,5	7,1
1002002	2019	2	7,5	6,5	7,0
1001003	2020	1	8,4	5,6	4,0
1002003	2020	1	7,0	4,5	9,0

Matrícula a procurar: 1001003



Matrícula	Ano	Semestre	P1	P2	P3
1001001	2019	1	10,0	8,0	7,0
1002001	2019	1	7,0	5,5	1,2
1003001	2019	1	4,0	7,0	8,0
1001002	2019	2	8,0	4,5	7,1
1002002	2019	2	7,5	6,5	7,0
1001003	2020	1	8,4	5,6	4,0
1002003	2020	1	7,0	4,5	9,0

Matrícula a procurar: 1001003

Matrícula	Ano	Semestre	P1	P2	P3
1001001	2019	1	10,0	8,0	7,0
1002001	2019	1	7,0	5,5	1,2
1003001	2019	1	4,0	7,0	8,0
1001002	2019	2	8,0	4,5	7,1
1002002	2019	2	7,5	6,5	7,0
1001003	2020	1	8,4	5,6	4,0
1002003	2020	1	7,0	4,5	9,0



Matrícula a procurar: 1001003



Matrícula	Ano	Semestre	P1	P2	P3
1001001	2019	1	10,0	8,0	7,0
1002001	2019	1	7,0	5,5	1,2
1003001	2019	1	4,0	7,0	8,0
1001002	2019	2	8,0	4,5	7,1
1002002	2019	2	7,5	6,5	7,0
1001003	2020	1	8,4	5,6	4,0
1002003	2020	1	7,0	4,5	9,0

Matrícula a procurar: 1001003

Matrícula	Ano	Semestre	P1	P2	P3
1001001	2019	1	10,0	8,0	7,0
1002001	2019	1	7,0	5,5	1,2
1003001	2019	1	4,0	7,0	8,0
1001002	2019	2	8,0	4,5	7,1
1002002	2019	2	7,5	6,5	7,0
1001003	2020	1	8,4	5,6	4,0
1002003	2020	1	7,0	4,5	9,0

Matrícula a procurar: 1001003

Matrícula	Ano	Semestre	P1	P2	P3
1001001	2019	1	10,0	8,0	7,0
1002001	2019	1	7,0	5,5	1,2
1003001	2019	1	4,0	7,0	8,0
1001002	2019	2	8,0	4,5	7,1
1002002	2019	2	7,5	6,5	7,0
1001003	2020	1	8,4	5,6	4,0
1002003	2020	1	7,0	4,5	9,0



Matrícula a procurar: 1001003

Matrícula	Ano	Semestre	P1	P2	P3
1001001	2019	1	10,0	8,0	7,0
1002001	2019	1	7,0	5,5	1,2
1003001	2019	1	4,0	7,0	8,0
1001002	2019	2	8,0	4,5	7,1
1002002	2019	2	7,5	6,5	7,0
1001003	2020	1	8,4	5,6	4,0
1002003	2020	1	7,0	4,5	9,0

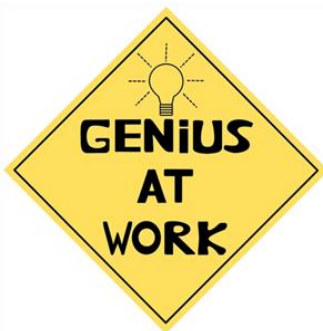
Cenário #2

Quero fazer análise estatística das notas

- Evolução da média das turmas ao longo dos anos

Proposta?

- implementar um programa que lê o arquivo sequencialmente
- calcula a média de cada turma
- imprime na tela os valores das médias por ano/semestre



Calcular a média das turmas



Matrícula	Ano	Semestre	P1	P2	P3	
1001001	2019	1	10,0	8,0	7,0	8,33
1002001	2019	1	7,0	5,5	1,2	
1003001	2019	1	4,0	7,0	8,0	
1001002	2019	2	8,0	4,5	7,1	
1002002	2019	2	7,5	6,5	7,0	
1001003	2020	1	8,4	5,6	4,0	
1002003	2020	1	7,0	4,5	9,0	

Calcular a média das turmas

Matrícula	Ano	Semestre	P1	P2	P3	
1001001	2019	1	10,0	8,0	7,0	8,33
1002001	2019	1	7,0	5,5	1,2	4,56
1003001	2019	1	4,0	7,0	8,0	
1001002	2019	2	8,0	4,5	7,1	
1002002	2019	2	7,5	6,5	7,0	
1001003	2020	1	8,4	5,6	4,0	
1002003	2020	1	7,0	4,5	9,0	



Calcular a média das turmas

Matrícula	Ano	Semestre	P1	P2	P3	
1001001	2019	1	10,0	8,0	7,0	8,33
1002001	2019	1	7,0	5,5	1,2	4,56
1003001	2019	1	4,0	7,0	8,0	6,33
1001002	2019	2	8,0	4,5	7,1	
1002002	2019	2	7,5	6,5	7,0	
1001003	2020	1	8,4	5,6	4,0	
1002003	2020	1	7,0	4,5	9,0	



Calcular a média das turmas

Matrícula	Ano	Semestre	P1	P2	P3	
1001001	2019	1	10,0	8,0	7,0	8,33
1002001	2019	1	7,0	5,5	1,2	4,56
1003001	2019	1	4,0	7,0	8,0	6,33
1001002	2019	2	8,0	4,5	7,1	
1002002	2019	2	7,5	6,5	7,0	
1001003	2020	1	8,4	5,6	4,0	
1002003	2020	1	7,0	4,5	9,0	



Calcular a média das turmas

Matrícula	Ano	Semestre	P1	P2	P3	
1001001	2019	1	10,0	8,0	7,0	8,33
1002001	2019	1	7,0	5,5	1,2	4,56
1003001	2019	1	4,0	7,0	8,0	6,33
1001002	2019	2	8,0	4,5	7,1	6,40
1002002	2019	2	7,5	6,5	7,0	
1001003	2020	1	8,4	5,6	4,0	
1002003	2020	1	7,0	4,5	9,0	

Média da turma 2019/1 = 6,40

Mas...

Busca por matrícula específica exige fazer uma leitura sequencial do arquivo

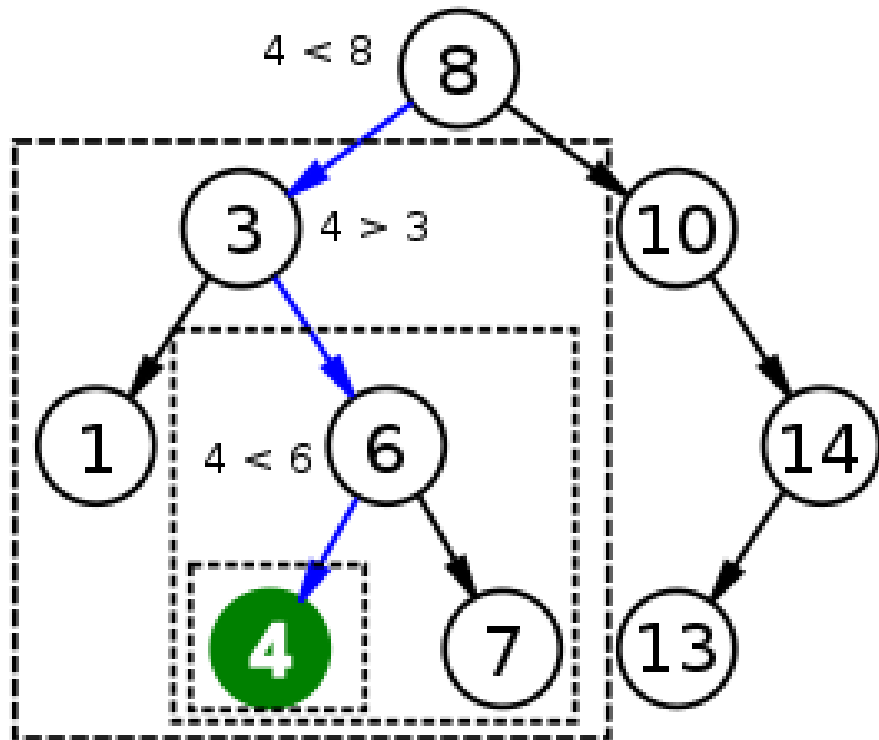
E quando o arquivo ficar muito grande?



Como melhorar esse cenário?

Ordenar por matrícula

Aplicar **busca binária**!



Matrícula a procurar: 1001003

Matrícula	Ano	Semestre	P1	P2	P3
1001001	2019	1	10	8	7
1001002	2019	2	8	4,5	7,1
1001003	2020	1	8,4	5,6	4
1002001	2019	1	7	5,5	1,2
1002002	2019	2	7,5	6,5	7
1002003	2020	1	7	4,5	9
1003001	2019	1	4	7	8



Matrícula a procurar: 1001003

Matrícula	Ano	Semestre	P1	P2	P3
1001001	2019	1	10	8	7
1001002	2019	2	8	4,5	7,1
1001003	2020	1	8,4	5,6	4
1002001	2019	1	7	5,5	1,2
1002002	2019	2	7,5	6,5	7
1002003	2020	1	7	4,5	9
1003001	2019	1	4	7	8

Matrícula a procurar: 1001003

Matrícula	Ano	Semestre	P1	P2	P3
1001001	2019	1	10	8	7
1001002	2019	2	8	4,5	7,1
1001003	2020	1	8,4	5,6	4
1002001	2019	1	7	5,5	1,2
1002002	2019	2	7,5	6,5	7
1002003	2020	1	7	4,5	9
1003001	2019	1	4	7	8

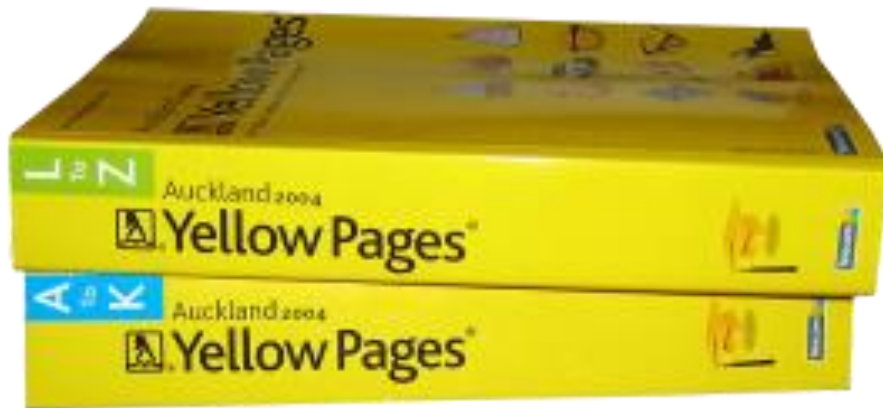
Por outro lado...

Calcular a média da turma ficou mais complexo
Os dados não estão mais organizados por turma!



Vamos então indexar!

Criar um arquivo separado que aponta para os registros que contêm um determinado valor de ano/semestre



Indexar por Ano/Semestre

	Matrícula	Ano	Sem.	P1	P2	P3
0	1001001	2019	1	10	8	7
1	1001002	2019	2	8	4,5	7,1
2	1001003	2020	1	8,4	5,6	4
3	1002001	2019	1	7	5,5	1,2
4	1002002	2019	2	7,5	6,5	7
5	1002003	2020	1	7	4,5	9
6	1003001	2019	1	4	7	8

Ano	Semestre	Endereço
2019	1	{0, 3, 6}
2019	2	{1, 4}
2020	1	{2, 5}

Mas...

O chefe do departamento **gostou da minha ideia**

Todos os professores agora devem usar meu programa!

- Adicionar uma coluna para identificar o professor da turma
- Adicionar uma coluna para identificar a disciplina
- Adicionar uma coluna para identificar a turma

Mudando meu arquivo...

Matrícula	Professor	Disciplina	Turma	Ano	Semestre	P1	P2	P3
1001001	Vanessa	BD2	A1	2019	1	10	8	7
1001001	Vanessa	Prog I	A1	2016	2	5	8,5	6,5
1001002	Daniel	BD2	A1	2019	2	8	4,5	7,1
1001002	Simone	FAC	A1	2016	2	4	6,5	7
1001003	Daniel	ED2	A1	2020	1	8,4	5,6	4
1002001	Vanessa	BD2	A1	2019	1	7	5,5	1,2
1002002	Daniel	BD2	A1	2019	2	7,5	6,5	7
1002003	Daniel	BD2	A1	2020	1	7	4,5	9
1003001	Vanessa	BD2	A1	2019	1	4	7	8

Alteração na estrutura do arquivo...

Implica em alteração **no programa** que lê o arquivo

Não existe **independência** entre os dados e o programa

Que outros problemas podem acontecer?

Mudando meu arquivo...

Matrícula	Professor	Disciplina	Turma	Ano	Semestre	P1	P2	P3
1001001	Vanessa	BD2	A1	2019	1	10	8	7
1001001	Vanessa	Prog I	A1	2016	2	5	8,5	6,5
1001002	Daniel	BD2	A1	2019	2	8	4,5	7,1
1001002	Simone	FAC	A1	2016	2	4	6,5	7
1001003	Daniel	ED2	A1	2020	1	8,4	5,6	4
1002001	Vanessa	BD2	A1	2019	1	7	5,5	1,2
1002002	Daniel	BD2	A1	2019	2	7,5	6,5	7
1002003	Daniel	BD2	A1	2020	1	7	4,5	9
1003001	Vanessa	BD2	A1	2019	1	4	7	8

Mudando meu arquivo...



Matrícula	Professor	Disciplina	Turma	Ano	Semestre	P1	P2	P3
1001001	Aline	BD2	A1	2019	1	10	8	7
1001001	Vanessa	Prog I	A1	2016	2	5	8,5	6,5
1001002	Daniel	BD2	A1	2019	2	8	4,5	7,1
1001002	Simone	FAC	A1	2016	2	4	6,5	7
1001003	Daniel	ED2	A1	2020	1	8,4	5,6	4
1002001	Vanessa	BD2	A1	2019	1	7	5,5	1,2
1002002	Daniel	BD2	A1	2019	2	7,5	6,5	7
1002003	Daniel	BD2	A1	2020	1	7	4,5	9
1003001	Vanessa	BD2	A1	2019	1	4	7	8

Mudando meu arquivo...

Matrícula	Professor	Disciplina	Ano	Semestre	P1	P2	P3
1001001	Aline	BD1	2019	1	10	8	7
1001001	Vanessa	BD1	2019	2	5	8,5	6,5
1001002	Daniel	BD1	2019	1	8	4,5	7,1
1001002	Simone	BD1	2019	2	4	6,5	7
1001003	Daniel	BD1	2019	1	8,4	5,6	4
1002001	Vanessa	BD1	2019	2	7	5,5	1,2
1002002	Daniel	BD1	2019	1	7,5	6,5	7
1002003	Daniel	BD1	2019	2	7	4,5	9
1003001	Vanessa	BD2	2019	1	4	7	8



Aprendizado #1

“Redundância leva a Inconsistência”

Acesso Concorrente

Eu quero inserir uma nova turma

- Abro o arquivo
- Gravo 20 novos registros

Ao mesmo tempo, Daniel vai inserir novos dados

- Abre o arquivo
- Grava 30 novos registros



iiiiihhhhhh....

Depende da ordem das operações

- Alguns dos meus registros são gravados
- Outros são sobrescritos pelos do Daniel
- Os 30 registros do Daniel sobrescrevem os meus 20...



Aprendizado #2

**“Acesso concorrente pode causar
inconsistência”**

Quem está vendo meus dados?

Matrícula	Professor	Disciplina	Turma	Ano	Semestre	P1	P2	P3
1001001	Vanessa	BD2	A1	2019	1	10	8	7
1001001	Vanessa	Prog I	A1	2016	2	5	8,5	6,5
1001002	Daniel	BD2	A1	2019	2	8	4,5	7,1
1001002	Simone	FAC	A1	2016	2	4	6,5	7
1001003	Daniel	ED2	A1	2020	1	8,4	5,6	4
1002001	Vanessa	BD2	A1	2019	1	7	5,5	1,2
1002002	Daniel	BD2	A1	2019	2	7,5	6,5	7
1002003	Daniel	BD2	A1	2020	1	7	4,5	9
1003001	Vanessa	BD2	A1	2019	1	4	7	8

Aprendizado #3

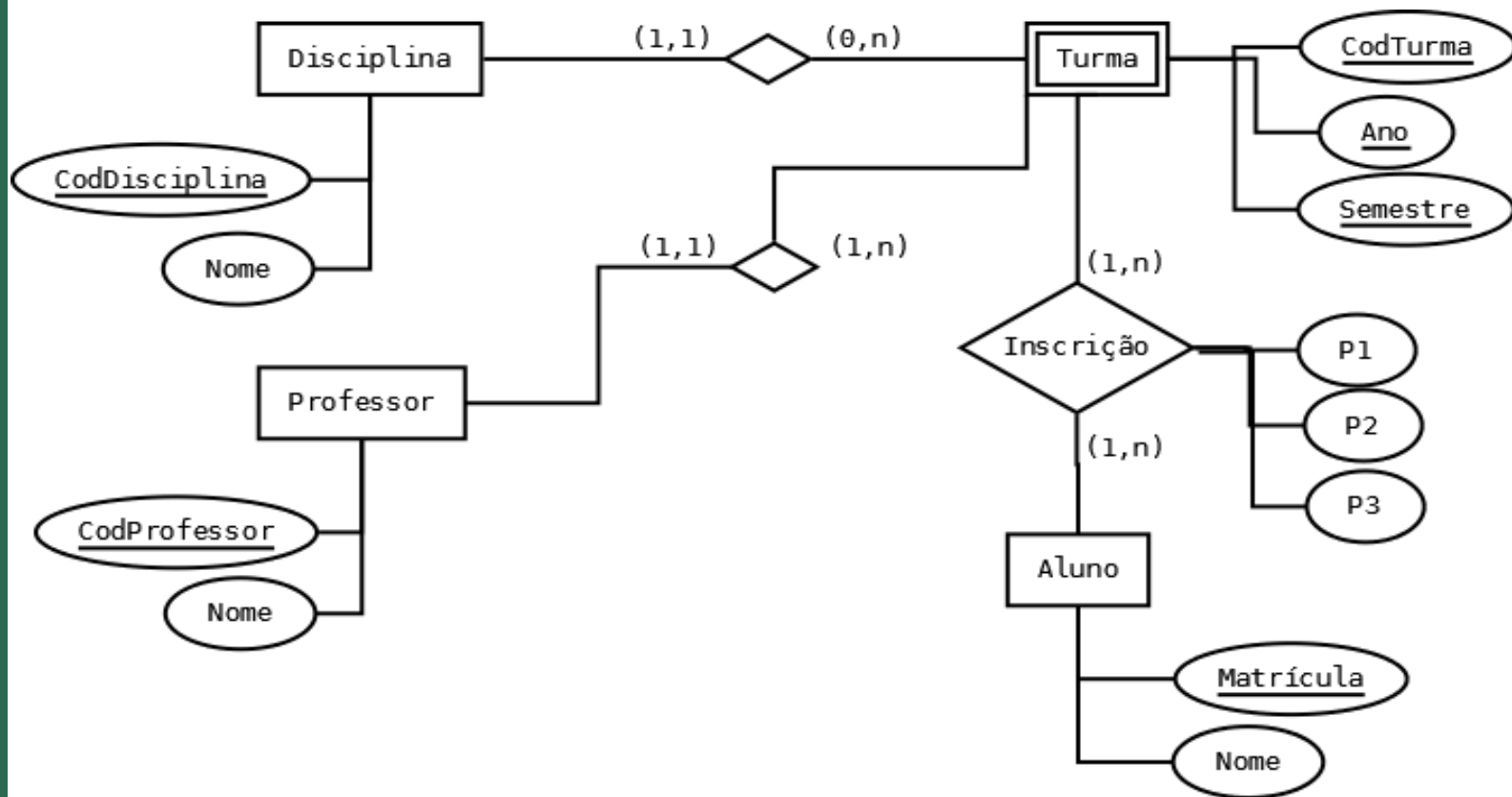
“Um usuário só pode ver seus próprios dados”

Como resolver esses problemas?

Uso de um Sistema de Gerência de Banco de Dados

- Controle de Concorrência
- Controle de Acesso
- Indexação
- Visões
- Triggers

Modelo Conceitual



Tradução para Modelo Relacional

Disciplina (CodDisciplina, Nome)

Professor (CodProfessor, Nome)

Turma (CodDisciplina, CodTurma, Ano, Semestre, CodProfessor)

- CodDisciplina referencia Disciplina
- CodProfessor referencia Professor

Aluno (Matricula, Nome)

Inscricao (Matricula, CodDisciplina, CodTurma, Ano, Semestre, P1, P2, P3)

- Matricula referencia Aluno
- (CodDisciplina, CodTurma, Ano, Semestre) referencia Turma

Criação das Tabelas Físicas

```
CREATE TABLE Disciplina (  
    codDisciplina VARCHAR(8) NOT NULL,  
    nome VARCHAR(100),  
    PRIMARY KEY (codDisciplina)  
);
```

```
CREATE TABLE Professor (  
    codProfessor VARCHAR(8) NOT NULL,  
    nome VARCHAR(100),  
    PRIMARY KEY (codProfessor)  
);
```



Aluno

matricula	nome
1001001	Adriana
1001002	Adriel
1001003	Bruno
1001004	Melissa
1001005	Joana
1001006	Julio
1001007	
1001008	
1001009	
1001010	
1001011	
1001012	
1001013	
1001014	Leandro
1001015	Carla
1001016	Julian
1001017	Maria
1001018	Gabriel
1001019	Pedro
1001020	Walter

Inscrição

matricula	codDisciplina	codTurma	ano	semestre	p1	p2	p3
1001001	TCC00173	A1	2012	1	7.8	1.8	2.9
1001002	TCC00215	A1	2012	1	9.2	7.2	5.9
1001003	TCC00173	A2	2011	2	8.9	2.7	0.6
1001004	TCC00215	A1	2012	1	1.6	2.7	9.6
1001005	TCC00173	A1	2010	2	9.6	7.8	2
1001006	TCC00215	A1	2011	2	1.4	1.1	7.5
1001007	TCC00173	A2	2009	2	9.4	9.2	4.8
1001008	TCC00215	A1	2011	2	3.6	9	5.4
1001009	TCC00173	A1	2009	2	0.2	6.6	5.7
1001010	TCC00215	A1	2011	2	1.5	5.2	1.2
1001011	TCC00173	A2	2008	1	0.3	5.1	5.5
1001012	TCC00215	A1	2010	1	3.4	8	2.8
1001013	TCC00173	A1	2012	1	3.3	3.4	7.3
1001014	TCC00215	A1	2010	1	5.9	8.4	5.4
1001015	TCC00173	A2	2011	2	4.8	9.1	1.5
1001016	TCC00215	A1	2009	2	4.8	7.6	4.7
1001017	TCC00173	A1	2010	1	8.4	9.8	0.5
1001018	TCC00215	A2	2009	2	2.7	5.4	1.5
1001019	TCC00173	A1	2008	1			

Turma

codDisciplina	codTurma	ano	semestre	codProfessor
TCC00215	A1	2008	1	500100
TCC00215	A1	2009	2	500100
TCC00215	A1	2010	1	500100
TCC00215	A1	2011	2	500100
TCC00215	A1	2012	1	500100
TCC00215	A2	2008	1	500100
TCC00215	A2	2009	2	500100
TCC00215	A2	2010	1	500100
TCC00215	A2	2011	2	500100
TCC00215	A2	2012	1	500100
TCC00165	A1	2008	1	500101
TCC00215	A1	2009	2	500101
TCC00215	A1	2010	1	500101
TCC00215	A1	2011	2	500101
TCC00215	A1	2012	1	500101
TCC00215	A2	2008	1	500101
TCC00215	A2	2009	2	500101
TCC00215	A2	2010	1	500101
TCC00215	A2	2011	2	500101
TCC00165	A2	2012	1	500101
TCC00173	A1	2008	1	500103
TCC00173	A1	2009	2	500103

1001002	Adriel
1001003	Bruno
1001004	Melissa
1001005	Joana
1001006	Julio
1001007	Marcelo

Disciplina

codDisciplina	nome
TCC00165	Fundamentos de Arquiteturas de Computadores
TCC00173	Programação de Computadores I
TCC00215	Estruturas de Dados II

Inscrição

matricula	codDisciplina	codTurma	ano	semestre	p1	p2	p3
1001001	TCC00173	A1	2012	1	7.8	1.8	2.9
1001001	TCC00215	A1	2012	1	9.2	7.2	5.9
1001002	TCC00173	A1	2011	2	8.9	2.7	0.6
1001002	TCC00215	A2	2012	1	1.6	2.7	0.5
1001003	TCC00173	A1	2010	1	9.6	7.8	9.6
1001003	TCC00215	A1	2011	2	1.4	7	2
1001004	TCC00173	A1	2009	2	9.4	1.1	7.5
1001004	TCC00215	A2	2011	2	3.6	9.2	4.8
1001005	TCC00173	A1	2008	1	0.2	9	5.4
1001005	TCC00215	A1	2010	1	1.5	6.6	5.7
1001006	TCC00173	A1	2012	1	0.3	5.2	1.2
1001006	TCC00215	A2	2010	1	3.4	8	5.1
1001007	TCC00173	A1	2011	2	3.3	3.4	2.8
1001007	TCC00215	A1	2009	2	5.9	8.4	7.3
1001008	TCC00173	A1	2010	1	4.8	9.1	5.4

Professor

codProfessor	nome
500100	Vanessa
500101	Simone
500103	Leo Murta

Turma

codDisciplina	codTurma	ano	semestre	codProfessor
TCC00215	A1	2008	1	500100
TCC00215	A1	2009	2	500100
TCC00215	A1	2010	1	500100
TCC00215	A1	2011	2	500100
TCC00215	A1	2012	1	500100
TCC00215	A2	2008	1	500100
TCC00215	A2	2009	2	500100
TCC00215	A2	2010	1	500100

Vamos consultar?

Listar todos os nomes de alunos que estão matriculados na turma “A1” da disciplina “TCC00165” em 2019/1.

```
SELECT nome
FROM aluno a, inscricao i
WHERE i.ano = 2019
AND i.semestre = 1
AND i.codTurma = 'A1'
AND i.codDisciplina = 'TCC00165'
AND a.matricula = i.matricula
```

Vamos consultar?

Listar todos os nomes de alunos que estão matriculados na turma “A1” da disciplina “TCC00165” em 2019/1.

```
SELECT nome
FROM aluno a, inscricao i
WHERE i.ano = 2019
AND i.semestre = 1
AND i.codTurma = 'A1'
AND i.codDisciplina = 'TCC00165'
AND a.matricula = i.matricula
```

nome
Daniel
Cristina

Apenas um exemplo....

2.93 bilhões de usuários

Mais de 8 bilhões de visualizações de vídeo por dia

317.000 atualizações de status por minuto



Instagram



Dados de 2024

Inserção de mais de **95 milhões** de fotos e vídeos por dia
Mais de **1 bilhão** de usuários

PostgreSQL



Dados de usuários,
fotos, tags



redis

Activity Feed



Cache
Distribuída

**Voltando à nossa pergunta: o
que é um banco de dados?**

Dado

**“fatos que podem ser armazenados,
e.g., nomes, telefones, endereços”**

Banco de Dados

“Coleção de dados inter-relacionados projetada para atender às necessidades de um grupo específico de usuários e de aplicações”

- Sinônimo: base de dados

Exemplo de Base de Dados

Componentes

- Alunos
- Professores
- Técnicos
- Disciplinas
- Turmas
- Departamentos
- Institutos



Relacionamentos

- Disciplinas são oferecidas por departamentos
- Departamentos fazem parte de institutos
- Alunos estão matriculados em turmas
- Turmas estão associadas a uma disciplina

Exemplo de Base de Dados

Componentes

- Alunos
- Professores
- Técnicos
- Disciplinas
- Turmas
- Departamentos
- Institutos



Relacionamentos

- Disciplinas **são oferecidas** por departamentos
- Departamentos **fazem parte** de institutos
- Alunos **estão matriculados** em turmas
- Turmas **estão associadas** a uma disciplina

Banco de Dados/Base de Dados

Representa um **modelo da realidade** (minimundo / universo de discurso)

É uma coleção de dados logicamente coerentes com algum significado inerente

Isolamento entre programas e dados

Contém a **definição** de sua própria **estrutura**

Contém **dados**, **metadados**, **índices** e dados sobre as aplicações que a utiliza

Abstração de Dados

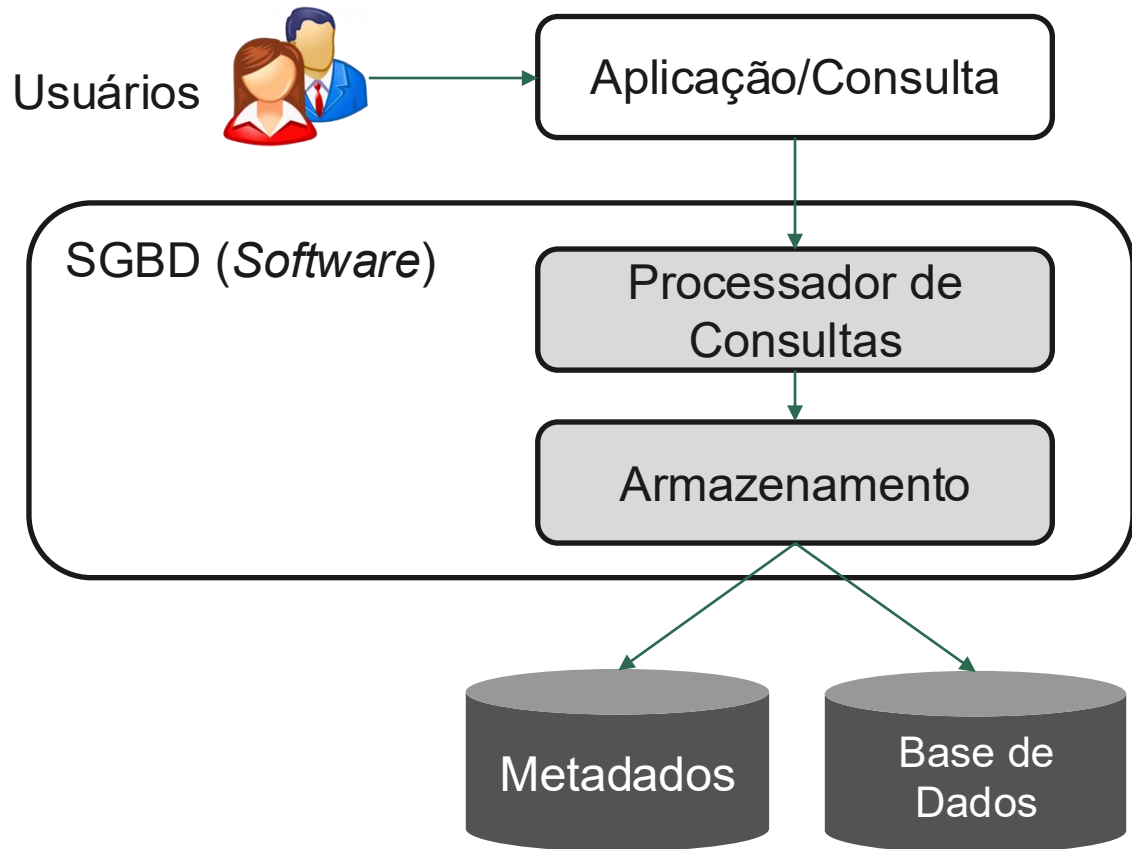


Sistema Gerenciador de Banco de Dados (SGBD)

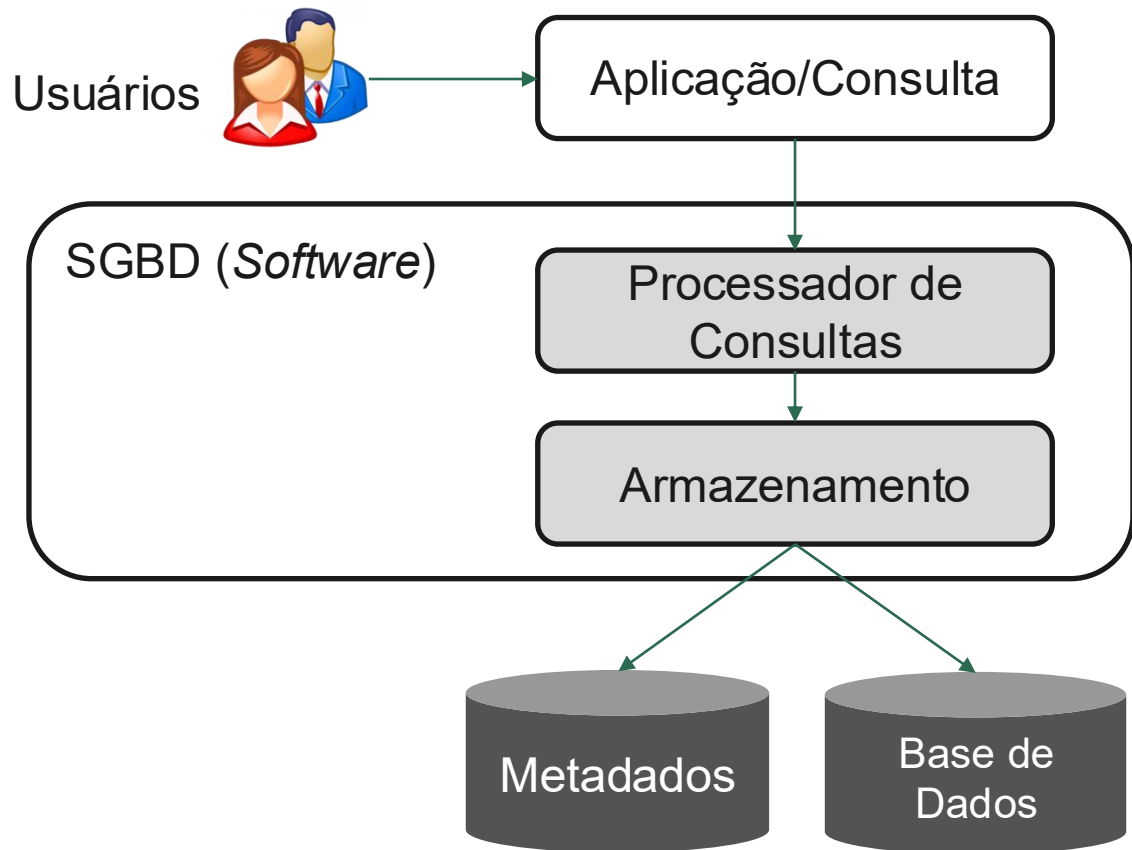
“Sistema (conjunto de programas) de propósito geral que facilita o processo de definição, construção e manipulação de bancos de dados para diversas aplicações”

- Exemplos:
 - Oracle, DB2, MySQL, PostgreSQL

SGBD



SGBD



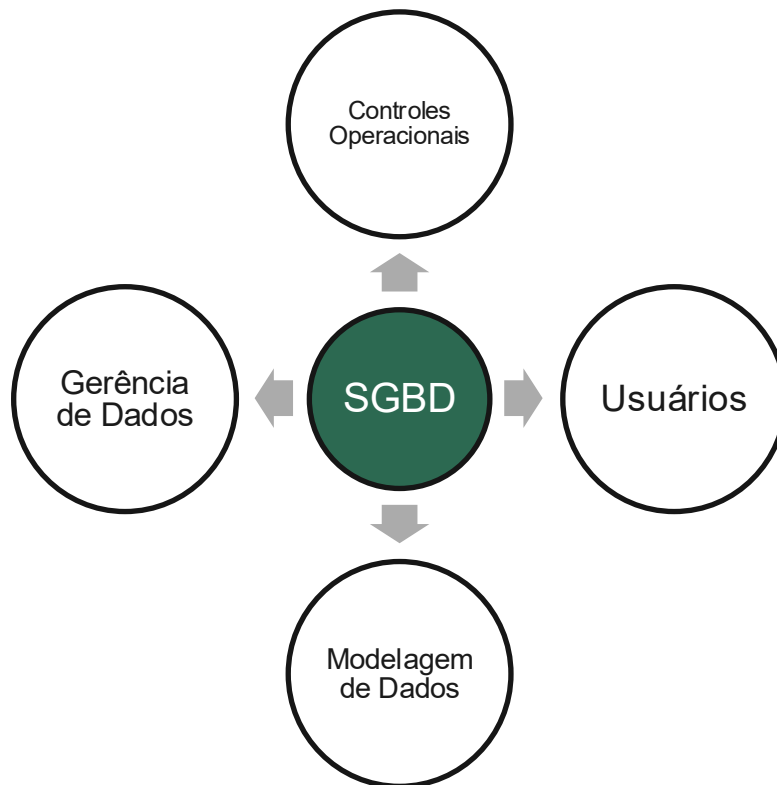
Funções do SGBD

Simplificar o desenvolvimento de aplicações caracterizadas por uso intensivo de **Dados**

Como?

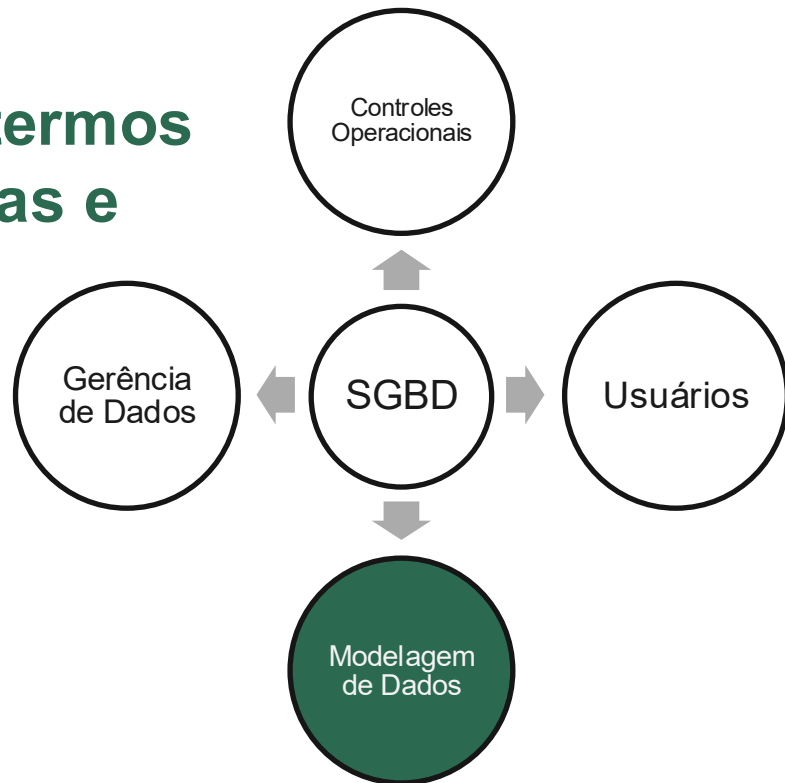
- Provendo serviços que **diminuem o tempo de desenvolvimento**
- Através de ferramentas o usuário pode:
 - Realizar **entrada de dados**
 - Examinar dados
 - **Manipular dados** de acordo com a aplicação

Funções do SGBD



Funções do SGBD

“Definir a base de dados em termos de tipos de dados, estruturas e restrições”



Funções do SGBD

**“Manipulação da Base de Dados:
consulta, geração de relatórios,
inserção, remoção e modificação do
seu conteúdo”**



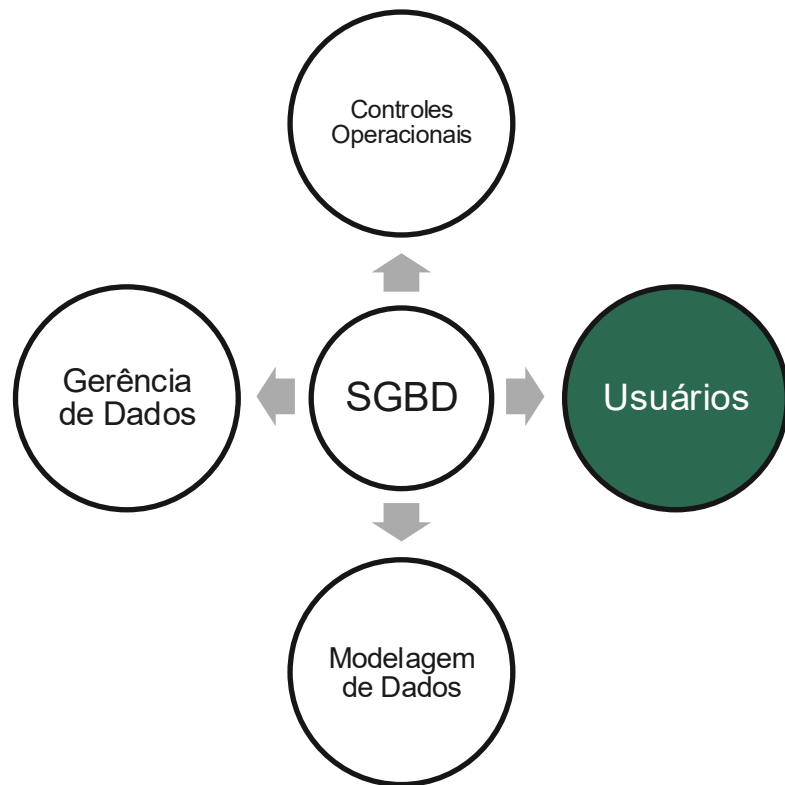
Funções do SGBD

“Processamento Concorrente e Compartilhamento de usuários e programas, porém mantendo todos os dados válidos e consistentes”



Funções do SGBD

“Projetistas da Aplicação e
Projetistas do SGBD”



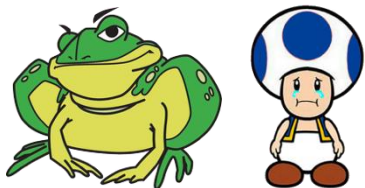
Tipos de Usuários

Projetista da Aplicação (quem aparece)

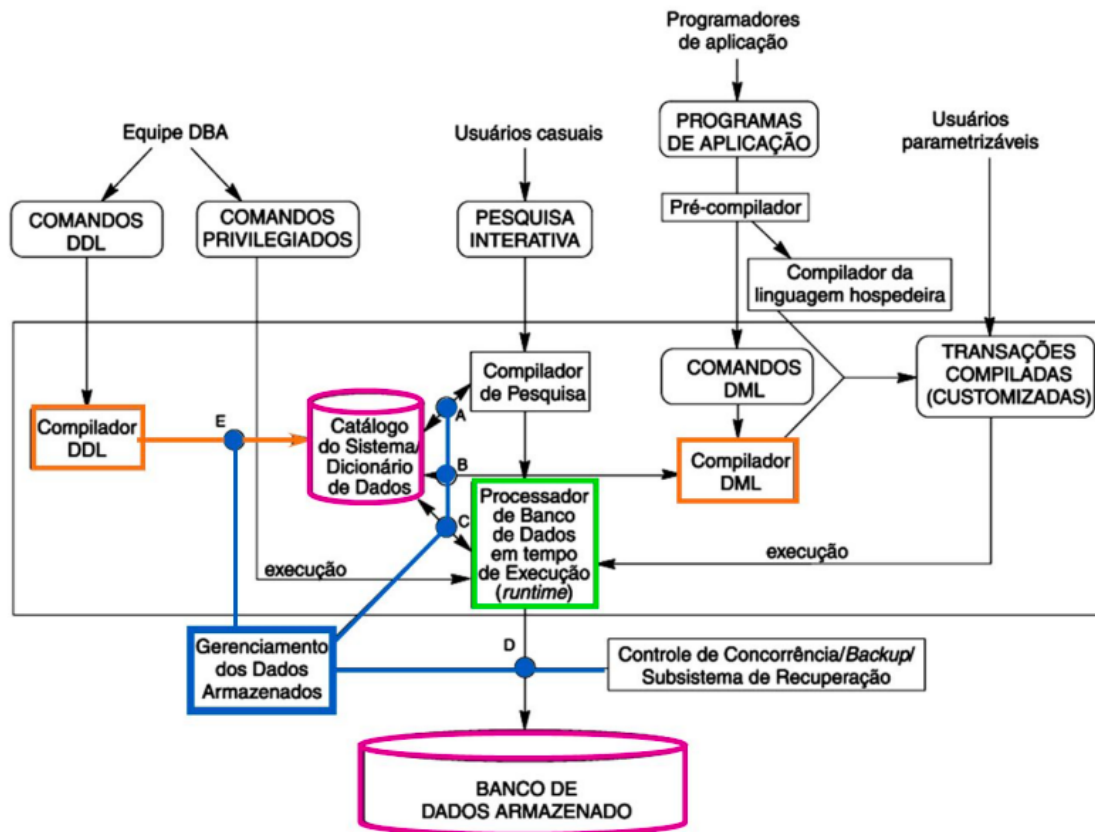
- Administrador de Banco de Dados
- Projetista da Base de Dados
- Usuário Final
- Programadores

Projetista do SGBD (quem fica no *backstage*)

- Projetistas dos componentes do SGBD
- Desenvolvedores de ferramentas (e.g., Toad)



SGBD Real...



Sistema de Banco de Dados

Aplicação

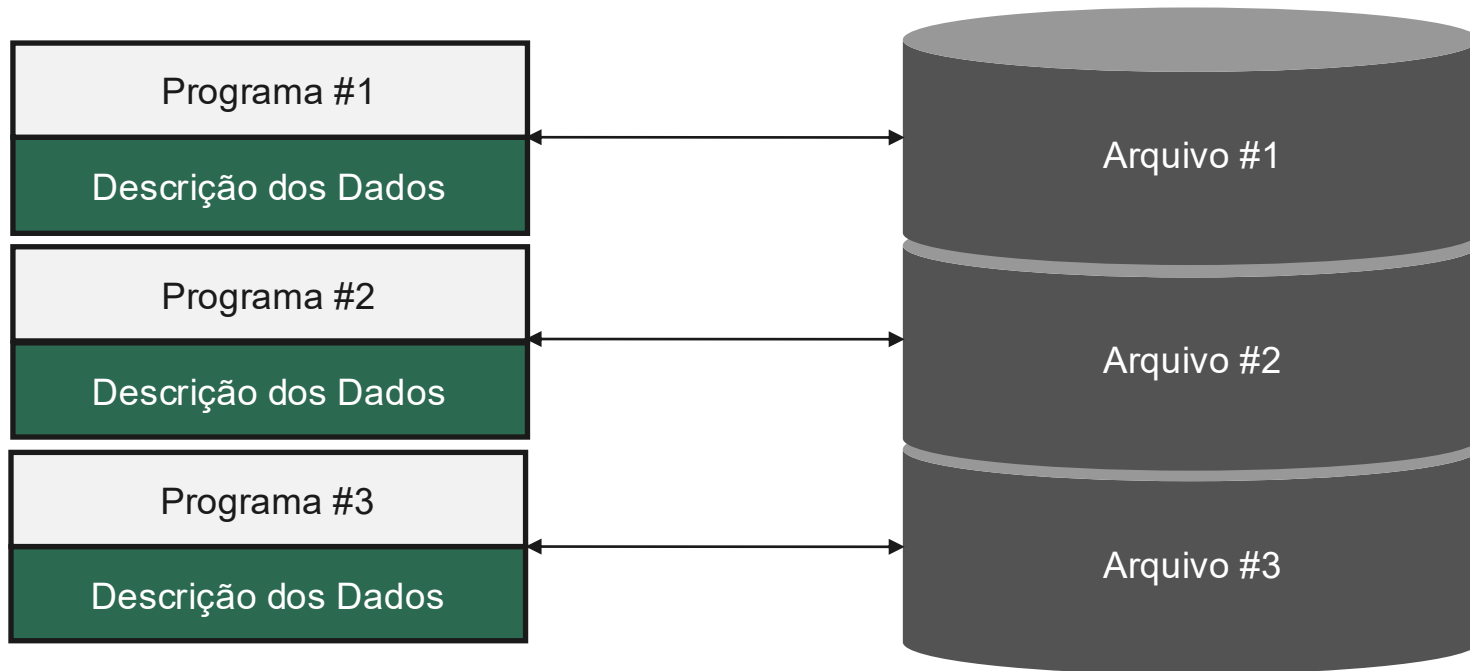
+

SGBD

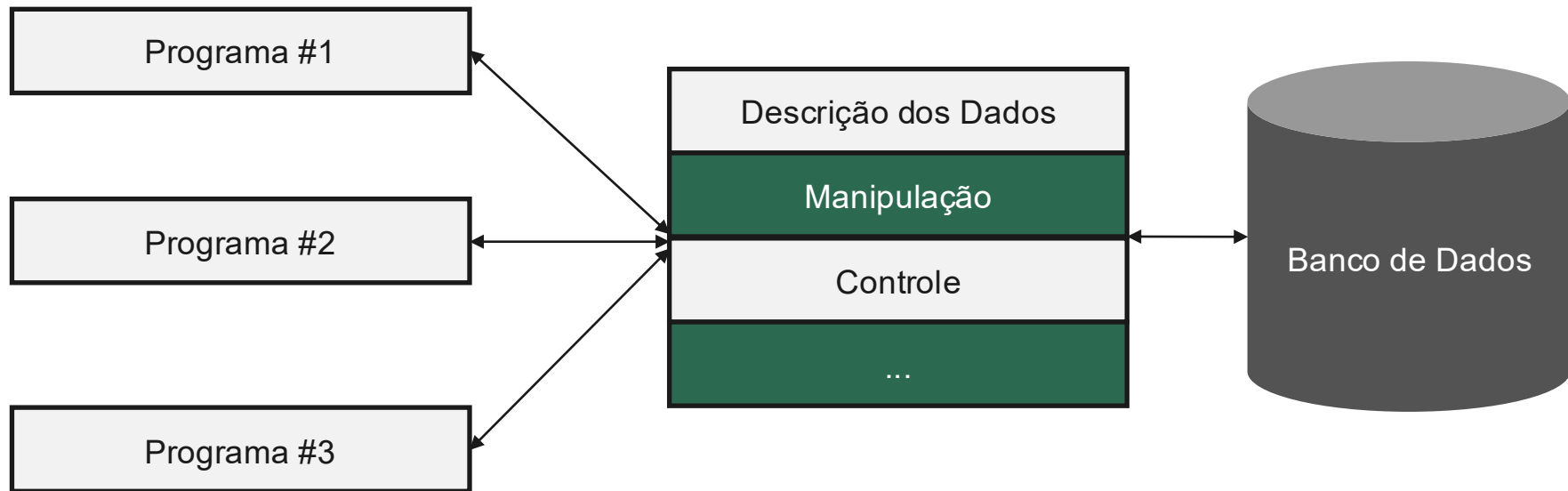
+

Base de Dados

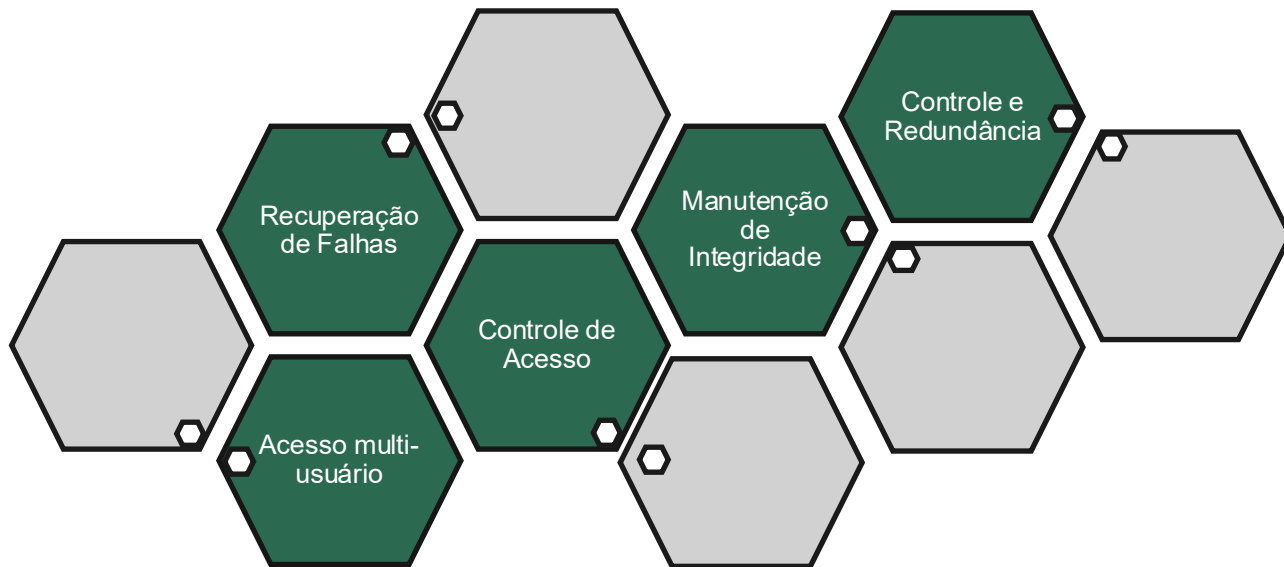
Sistema de Arquivos



Sistema de Banco de Dados



Vantagens da Utilização de BDs



Quando não usar Banco de Dados

Aplicações simples e bem definidas, sem mudanças

Aplicações (algumas) de tempo real

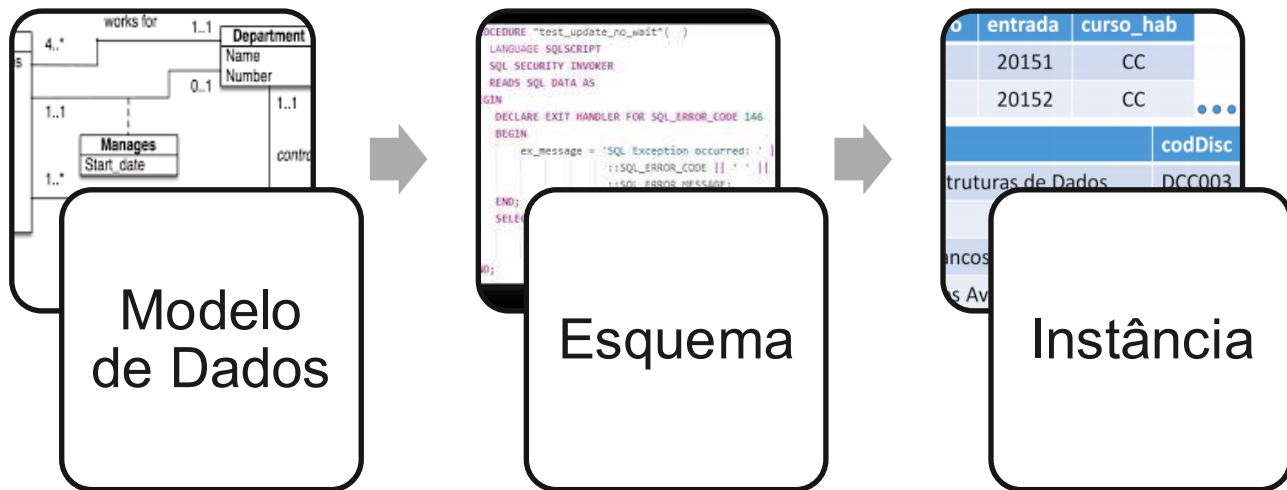
Aplicações sem acesso multiusuário

Motivos:

- Investimento inicial alto
- Generalidade na definição e manipulação dos dados
- Custo adicional para prover outras facilidades funcionais (manutenção de segurança, controle de concorrência, recuperação de falhas, etc.)

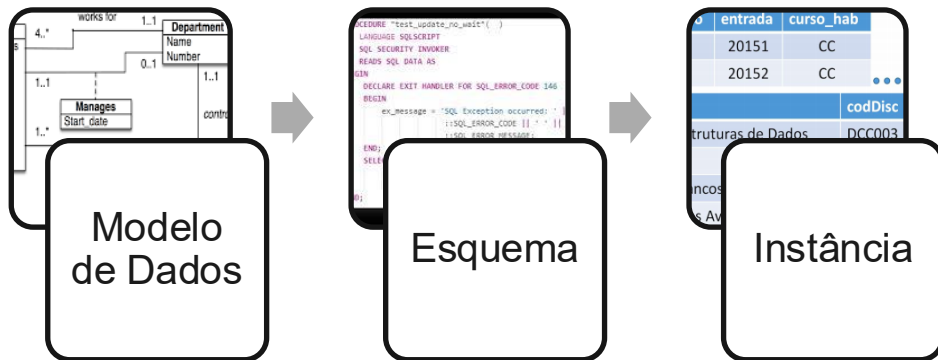


Mais conceitos importantes...



Mais conceitos importantes...

IMPORTANTE



Modelo de dados

- Conjunto de conceitos usados para descrever a estrutura de um BD
- Abstração de dados
- Estrutura = tipos de dados + relacionamentos + restrições (+ operações)
- Exemplos: Modelo Entidade-Relacionamento, Modelo Relacional

Esquema

- Descrição (textual ou gráfica) da estrutura de um BD

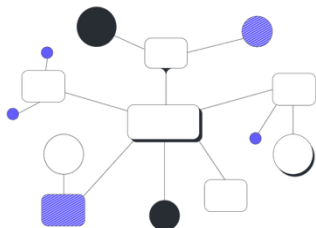
Instância

- Conjunto de dados armazenados em um BD em um determinado instante de tempo

Categorias de Modelos

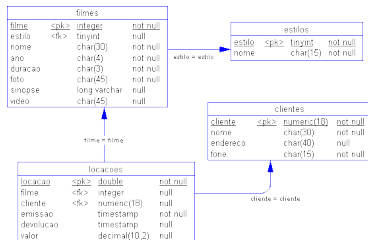
Conceitual

Descrevem a estrutura do banco de dados de uma forma próxima da percepção do usuários



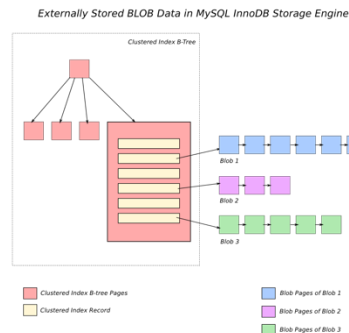
Lógico

Descrevem a estrutura do banco de dados de uma forma próxima à que será manipulada pelo SGBD



Físico

Descrevem como os dados são fisicamente armazenados



Bibliografia

Fundamentals of Database Systems, Elmasri e Navathe, caps. 1 e 2

Agradecimentos

Material de Daniel de Oliveira e Alexandre Plastino

Perguntas?

Vanessa Braganholo

vanessa@ic.uff.br