

MBA em IA e Big Data

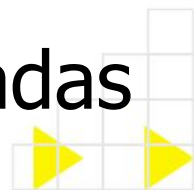


Curso 01 - Linguagens e Ferramentas para Inteligência Artificial e Big Data (Python e SQL)

Consultas avançadas com SQL - Agregações

Jose Fernando Rodrigues Junior
ICMC-USP São Carlos

Objetivo: prática com conceitos de junção, agregação, ordenação, operações com conjuntos, e consultas aninhadas



SELECT

- Funções Agregadas

- entrada \Rightarrow conjunto de valores

- saída \Rightarrow 1 valor

- Exemplos:

- **AVG(*atributo*)** \Rightarrow calcula a média da coluna *atributo*

- **COUNT()**

- **count(*)** – retorna o número de tuplas de uma consulta

- **count(*atributo*)** – retorna o nro de valores (com repetição) da coluna *atributo* que não tem valores null



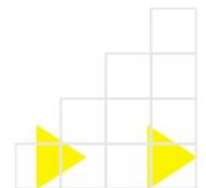
SELECT

- Funções Agregadas

- Exemplos

- **MAX**(*atributo*) ⇒ recupera o valor máximo da coluna *atributo*
 - **MIN**(*atributo*) ⇒ recupera o valor mínimo da coluna *atributo*
 - **SUM**(*atributo*) ⇒ obtém a soma de valores da coluna *atributo*
 - **AVG** (média), **STDDEV** (desvio padrão), e **VARIANCE** (variância)

<https://docs.oracle.com/database/121/SQLRF/functions003.htm#SQLRF20035>





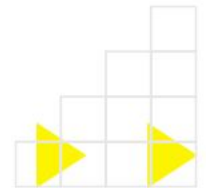
SELECT

- Funções Agregadas

- Exemplos

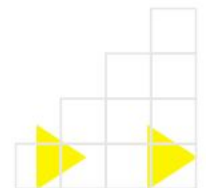
- **GROUP BY**, ou agrupamento, assume a presença de valores repetidos \Rightarrow portanto, apesar de possível, não faz sentido a realização de agrupamentos sobre (by) os atributos chave
 - **MAX** (
 - **MIN** (
 - **SUM** (
 - *AVG* (média), *STDDEV* (desvio padrão), e *VARIANCE* (variância)

<https://docs.oracle.com/database/121/SQLRF/functions003.htm#SQLRF20035>



SELECT

- **GROUP BY** \Rightarrow agrupamento de tuplas
 - para a aplicação de funções agregadas
- **HAVING** \Rightarrow condições aplicadas **a grupos já formados** por **GROUP BY**
- **ORDER BY** \Rightarrow estabelece a ordenação lógica da tabela de resultados
 - **ASC** (*default*)
 - **DESC**



Exemplo:

Aluno = {Nome, NUSP}

{<Zeca, 11111>,
<Zico, 22222>,
<Juca, 33333>,
<Tuca, 44444> }

Matricula= {Sigla, Numero, Aluno, Ano, Nota}

{<SCC-125, 1, 11111, 2010, 5.0>,
<SCC-148, 1, 11111, 2010, 7.0>,
<SCC-125, 2, 22222, 2010, 5.0>,
<SCC-148, 1, 22222, 2009, 4.0>}

- Selecionar, para cada aluno, seu nome e a média das notas das disciplinas em que foi aprovado (nota ≥ 5). Ordenar por nome de aluno

1º Passo: seleção e junção

```
SELECT ...  
  FROM Aluno A JOIN Matricula M  
           ON M.Aluno = A.NUSP  
 WHERE M.Nota BETWEEN 5.0 AND 10.0
```

{Nome, NUSP, Sigla, Nota}
{<Zeca, 11111, SCC-125, 5.0>,
<Zeca, 11111, SCC-148, 7.0>,
<Zico, 22222, SCC-125, 5.0>}

Exemplo: (continuação)

2º Passo: agrupamento e agregação

```
SELECT A.Nome, AVG(M.Nota) as Media
FROM Aluno A JOIN Matricula M
      ON M.Aluno = A.NUSP
WHERE M.Nota BETWEEN 5.0 AND 10.0
GROUP BY A.Nome
ORDER BY A.Nome;
```

Grupo Zeca

<SCC125, 5.0>
<SCC148, 7.0>

Grupo Zico

<SCC125, 5.0>

Função **AVG** aplicada sobre cada grupo



{Nome, Media}
{<Zeca, 6.0>,
<Zico, 5.0>}

Exemplo:

Aluno = {Nome, NUSP}

{<Zeca, 11111>,
<Zico, 22222>,
<Juca, 33333>,
<Tuca, 44444> }

Matricula= {Sigla, Numero, Aluno, Ano, Nota}

{<SCC-541, 1, 11111, 2009, 3.0>,
<SCC-541, 1, 11111, 2010, 7.0>,
<SCC-240, 1, 11111, 2010, 5.0>,
<SCC-240, 1, 22222, 2009, 4.0>}

Disciplina = {Sigla, Nome}

{<SCC-541, LabBD>,
<SCC-240, BD>}

- Selecionar os nomes dos alunos que fizeram uma mesma disciplina mais de uma vez. Listar também o nome da disciplina, o nro de vezes que cursou e a nota máxima que o aluno obteve (considerando todas as vezes que cursou).

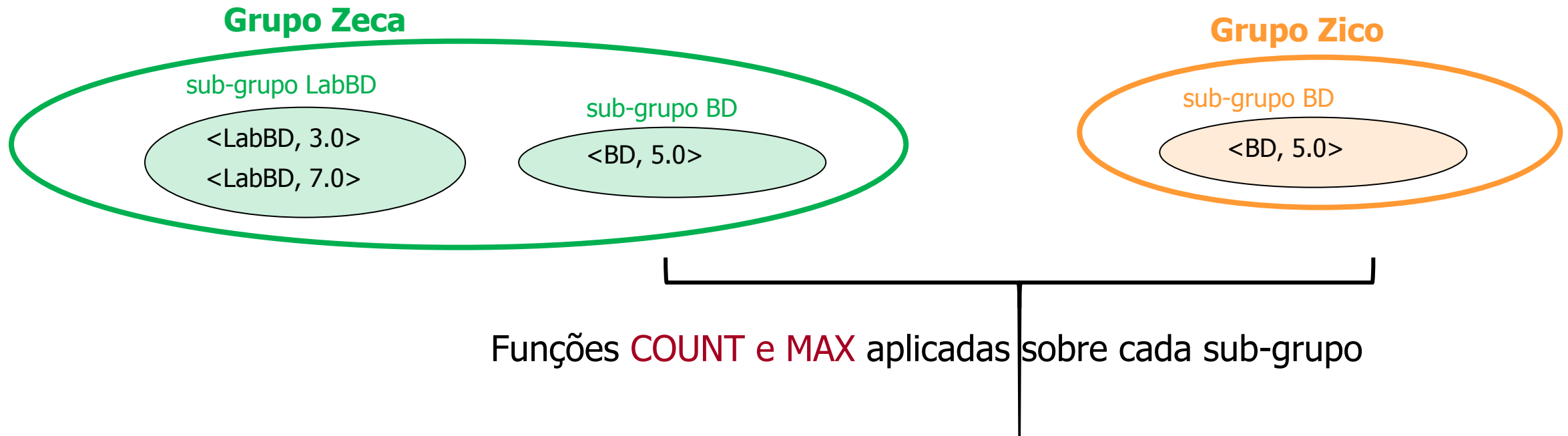
1º Passo: junção

```
select ....  
  from Aluno A join Matricula M  
           on A.NUSP = M.Aluno  
         join Disciplina D  
           on D.Sigla = M.Sigla
```


Exemplo: (continuação)

2º Passo: agrupamento e agregação

```
select A.Nome, D.Nome, count(*), max(M.Nota)
  from Aluno A join Matricula M
              on A.NUSP = M.Aluno
      join Disciplina D
      on D.Sigla = M.Sigla
 group by A.Nome, D.Nome
```



Exemplo: (continuação)

3º Passo: condição having

```
select A.Nome, D.Nome, count(*), max(M.Nota)
  from Aluno A join Matricula M
           on A.NUSP = M.Aluno
       join Disciplina D
           on D.Sigla = M.Sigla
 group by A.Nome, D.Nome
 having count(*) > 1
```

Grupo Zeca

sub-grupo LabBD

<LabBD, 3.0>

<LabBD, 7.0>

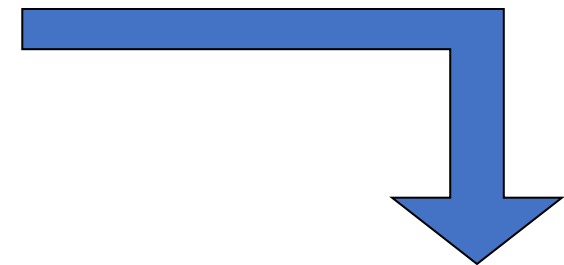
sub-grupo BD

<BD, 5.0>

Grupo Zico

sub-grupo BD

<BD, 5.0>



{A.Nome, D.Nome, count, max}
{< Zeca, LabBD, 2, 7.0>}