

SQL: Agregações

Fernando Lobo

Base de Dados, Universidade do Algarve

Agregações

- os operadores SUM, MIN, MAX, AVG, e COUNT, podem ser aplicados a um atributo de uma tabela.
- os operadores aparecem na cláusula SELECT. O resultado é um valor agregado para o respectivo atributo.
- COUNT(*) conta o número de tuplos.

Exemplos

- Qual a duração média dos filmes da Disney?

```
SELECT AVG(duracao)
FROM Filmes
WHERE nomeEstudio = 'Disney';
```

$$\frac{\text{avg}}{113.667}$$

- Quantos actores existem na tabela de Actores?

```
SELECT COUNT(*) FROM Actores;
```

$$\frac{\text{count}}{18}$$

NULLs são ignorados

- um valor NULL não conta para uma soma, média ou contagem.
- um valor NULL nunca pode ser o mínimo ou máximo de uma coluna.
- mas se todos os valores de uma coluna forem NULL, então o resultado da agregação é NULL.

Exemplo

```
--  
-- número de filmes da Disney  
--  
SELECT COUNT(*)  
    FROM Filmes  
    WHERE nomeEstudio = 'Disney';
```

```
--  
-- número de filmes da Disney  
-- cuja duração é conhecida  
--  
SELECT COUNT(duracao)  
    FROM Filmes  
    WHERE nomeEstudio = 'Disney';
```

Agrupar

- após um SELECT-FROM-WHERE podemos utilizar GROUP BY seguido de uma lista de atributos.
- o resultado do SELECT-FROM-WHERE é agrupado de acordo com os valores desses atributos.
- as agregações são aplicadas a cada grupo.

Exemplo

- Quantos filmes é que cada estúdio já produziu?

```
SELECT nomeEstudio, COUNT(*)  
      FROM Filmes  
GROUP BY nomeEstudio;
```

nomeEstudio	count
Disney	3
Fox	8
Paramount	5
Universal	5
Warner Brothers	1

Como funciona?

nome	ano	duracao	aCores	nomeEstudio
Mighty Ducks	1991	104	t	Disney
Lion King	1994	122	t	Disney
Pocahontas	1995	115	t	Disney
Star Wars	1977	124	t	Fox
Empire Strikes Back	1980	143	t	Fox
Return of the Jedi	1983	165	t	Fox
Total Recall	1990	110	t	Fox
Top Gun	1986	145	t	Fox
Moulin Rouge	2001	124	t	Fox
JFK	1996	130	t	Fox
Dances with Wolves	1990	132	t	Fox
Wayne's World	1992	95	t	Paramount
Addams Family	1991	102	t	Paramount
Gone With the Wind	1939	181	f	Paramount
Basic Instinct	1986	100	t	Paramount
Eyes Wide Shut	1999	148	t	Paramount
King Kong	1933		f	Universal
King Kong	1976	130	t	Universal
Indiana Jones	1981	130	t	Universal
The Fugitive	1993	120	t	Universal
Bridges of Madison County	1995	147	t	Universal
Batman Returns	1992	122	t	Warner Brothers

- COUNT(*) é aplicado a cada grupo.

Outro exemplo

- Quantos filmes é que cada atriz fez?

```
SELECT nomeActor AS 'atriz', COUNT(*)  
  FROM Participa, Actores  
 WHERE nomeActor = nome  
       AND sexo='f'  
 GROUP BY nomeActor;
```

atriz	count
Carrie Fisher	3
Dana Carvey	1
Meryl Streep	1
Michelle Pfeiffer	1
Nicole Kidman	2
Sharon Stone	2

Mais um exemplo

- Qual a duração máxima e mínima de cada estúdio?

```
SELECT nomeEstudio AS 'estudio',  
       MAX(duracao), MIN(duracao)  
FROM Filmes  
GROUP BY nomeEstudio;
```

estudio	max	min
Disney	122	104
Fox	165	110
Paramount	181	95
Universal	147	120
Warner Brothers	122	122

Restrições nos elementos a especificar em SELECT

- se usarmos um operador de agregação, então cada elemento especificado em SELECT tem de ser:
 - ① uma agregação, ou
 - ② um atributo especificado em GROUP BY.

Exemplo de um query inválido

- Qual a filme da Disney com maior duração?

```
-- incorrecto
SELECT nome, MAX(duracao)
  FROM Filmes
 WHERE nomeEstudio='Disney';
```

- a query é inválida.
- nome não é uma agregação, e também não aparece especificado em GROUP BY.

Porque é que é inválido?

```
-- incorrecto
SELECT nome, MAX(duracao)
  FROM Filmes
 WHERE nomeEstudio='Disney';
```

nome	ano	duracao	nomeEstudio	...
Mighty Ducks	1991	104	Disney	...
Lion King	1994	122	Disney	...
Pocahontas	1995	115	Disney	...

- `MAX(duracao)` → 122.
nome → ???
- Não faz sentido seleccionar nome.

Cláusula HAVING

- podemos ter HAVING <condição> a seguir a uma cláusula GROUP BY.
- a condição é aplicada a cada grupo.
- os grupos que satisfazem a condição vão para o output, os outros são eliminados.

Exemplo

- Quais as atrizes que fizeram mais do que 1 filme?

```
SELECT nomeActor AS 'atriz', COUNT(*)
  FROM Participa, Actores
 WHERE nomeActor = nome
       AND sexo='f'
 GROUP BY nomeActor
 HAVING COUNT(*) > 1;
```

atriz	count
Carrie Fisher	3
Nicole Kidman	2
Sharon Stone	2