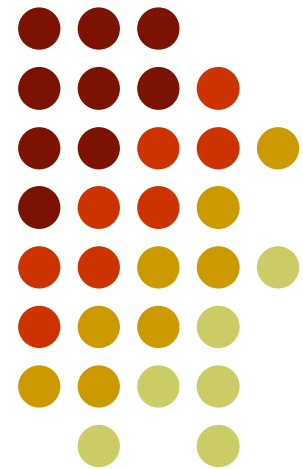


SQL: Linguaggio di query – funzioni di gruppo

Elena Ferrari
Basi di Dati
A.A. 2020/2021



Funzioni di gruppo



- Nelle espressioni della clausola di proiezione possiamo avere anche espressioni che calcolano valori a partire da insiemi di tuple:
 - media degli stipendi
 - massimo voto conseguito



Funzioni di gruppo

- Operano su un insieme di valori
- Producono come risultato un unico valore (aggregato)
- Sono indicate nella clausola di **SELECT**
 - Possono essere indicate più funzioni di gruppo contemporaneamente



Funzioni di gruppo

- Principali funzioni di gruppo previste da SQL:
 - **MAX**: determina il massimo in un insieme di valori
 - **MIN**: determina il minimo in un insieme di valori
 - **SUM**: esegue la somma dei valori in un insieme
 - **AVG**: esegue la media dei valori in un insieme
 - **COUNT**: determina la cardinalità di un insieme

Esempio



Film

titolo	regista	anno	genere	valutaz
underground	emir kusturica	1995	drammatico	3.20
edward mani di forbice	tim burton	1990	fantastico	3.60
nightmare before christmas	tim burton	1993	animazione	4.00
ed wood	tim burton	1994	drammatico	4.00
mars attacks	tim burton	1996	fantascienza	3.00
il mistero di sleepy hollow	tim burton	1999	horror	3.50
big fish	tim burton	2003	fantastico	3.10
la sposa cadavere	tim burton	2005	animazione	3.50
la fabbrica di cioccolato	tim burton	2005	fantastico	4.00
io non ho paura	gabriele salvatores	2003	drammatico	3.50
nirvana	gabriele salvatores	1997	fantascienza	3.00
mediterraneo	gabriele salvatores	1991	commedia	3.80
pulp fiction	quentin tarantino	1994	thriller	3.50
le iene	quentin tarantino	1992	thriller	4.00



Funzioni di gruppo

```
SELECT MIN(valutaz), AVG(valutaz), MAX(valutaz)
FROM Film;
```

```
SELECT MIN(valutaz), AVG(valutaz), MAX(valutaz)
FROM Film
WHERE genere = 'drammatico';
```



Funzioni di gruppo

- Ad accezione di COUNT, sono applicate solo su insiemi di valori semplici e non su insiemi di tuple
- L'insieme di valori è denotato nel caso più semplice da un nome di colonna, in generale da un'espressione contenente nomi di colonna
- Le funzioni SUM e AVG sono definite solo per insiemi di valori numerici



Funzione COUNT

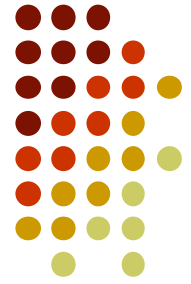
- Può avere come argomento anche il carattere speciale '*' nel qual caso la funzione restituisce il numero di tuple

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




Funzioni di gruppo

- Tutte le funzioni di gruppo possono essere usate con il qualificatore **DISTINCT**:
 - eventuali valori duplicati sono eliminati prima di applicare la funzione
 - l'eliminazione dei duplicati è significativa solo per le funzioni SUM, AVG e COUNT
- I valori nulli vengono eliminati prima del calcolo della funzione di gruppo



Funzioni di gruppo

- Se l'insieme dei valori cui applicare la funzione di gruppo è vuoto:
 - La funzione COUNT restituisce il valore numerico 0
 - Le altre funzioni restituiscono il valore NULL
- Le funzioni di gruppo possono a loro volta essere utilizzate in espressioni aritmetiche:
es. `SUM(valutaz)/COUNT(*)`

COUNT DISTINCT



```
SELECT COUNT(Reddito) FROM Persone;
```

```
SELECT COUNT(DISTINCT Reddito) FROM Persone;
```



Maternità

Madre	Figlio
Luisa	Maria
Luisa	Luigi
Anna	Olga
Anna	Filippo
Maria	Andrea
Maria	Aldo

Paternità

Padre	Figlio
Sergio	Franco
Luigi	Olga
Luigi	Filippo
Franco	Andrea
Franco	Aldo

Persone

Nome	Età	Reddito
Andrea	27	21
Aldo	25	15
Maria	55	42
Anna	50	35
Filippo	26	30
Luigi	50	40
Franco	60	20
Olga	30	41
Sergio	85	35
Luisa	75	87



Funzioni di gruppo: COUNT

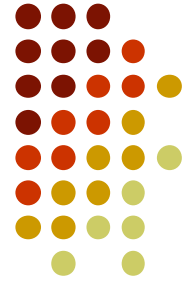
- Il numero di figli di Franco:

```
SELECT COUNT(*) AS NumFigliDiFranco  
FROM Paternita  
WHERE Padre = 'Franco';
```
- La funzione di gruppo (count) viene applicata al risultato dell'interrogazione:

```
SELECT *  
FROM Paternita  
WHERE Padre = 'Franco';
```

Paternità

Padre	Figlio
Sergio	Franco
Luigi	Olga
Luigi	Filippo
Franco	Andrea
Franco	Aldo



NumFigliDiFranco

2

Funzioni di gruppo



- Le colonne ottenute dall'applicazione di funzioni di gruppo, come quelle ottenute dall'applicazione di funzioni aritmetiche, sono colonne virtuali, cui è possibile assegnare un nome con la clausola AS



Funzioni di gruppo

- Le funzioni di gruppo sono valutate solo dopo l'applicazione di tutti i predicati nella clausola WHERE
- le funzioni di gruppo non possono essere utilizzate nella clausola WHERE (se non in sottointerrogazioni)



Funzioni di gruppo

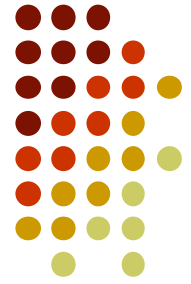
- Determinare il titolo e il regista del film drammatico di valutazione minima:

~~SELECT titolo, regista~~

~~FROM Film~~

~~WHERE genere = 'drammatico' AND valutaz = MIN(valutaz);~~

- Avremo bisogno di una sotto-interrogazione (vedi seguito)



Funzioni di gruppo

- Un'interrogazione scorretta:

```
SELECT Nome, MAX(Reddito)  
FROM Persone;
```
- Di chi sarebbe il nome? La clausola di proiezione deve essere omogenea:

```
SELECT MIN(Età), AVG(Reddito)  
FROM Persone;
```



Raggruppamento

- Le funzioni di gruppo possono essere applicate **a partizioni delle relazioni**
- Le colonne da usare nel partizionamento sono specificate nella clausola **GROUP BY**

Esempio



Film

titolo	regista	anno	genere	valutaz
underground	emir kusturica	1995	drammatico	3.20
edward mani di forbice	tim burton	1990	fantastico	3.60
nightmare before christmas	tim burton	1993	animazione	4.00
ed wood	tim burton	1994	drammatico	4.00
mars attacks	tim burton	1996	fantascienza	3.00
il mistero di sleepy hollow	tim burton	1999	horror	3.50
big fish	tim burton	2003	fantastico	3.10
la sposa cadavere	tim burton	2005	animazione	3.50
la fabbrica di cioccolato	tim burton	2005	fantastico	4.00
io non ho paura	gabriele salvatores	2003	drammatico	3.50
nirvana	gabriele salvatores	1997	fantascienza	3.00
mediterraneo	gabriele salvatores	1991	commedia	3.80
pulp fiction	quentin tarantino	1994	thriller	3.50
le iene	quentin tarantino	1992	thriller	4.00

Esempio



- Per ogni regista vogliamo determinare quanti suoi film sono presenti in catalogo, di quanti generi diversi e la valutazione minima, media e massima di tali film:

```
SELECT regista,  
       COUNT(*) AS numF,  
       COUNT(DISTINCT genere) AS numG,  
       MIN(valutaz) AS minV,  
       AVG(valutaz) AS avgV,  
       MAX(valutaz) AS maxV  
FROM Film  
GROUP BY regista;
```

Esempio



Film

titolo	regista	anno	genere	valutaz
underground	emir kusturica	1995	drammatico	3.20
edward mani di forbice	tim burton	1990	fantastico	3.60
nightmare before christmas	tim burton	1993	animazione	4.00
ed wood	tim burton	1994	drammatico	4.00
mars attacks	tim burton	1996	fantascienza	3.00
il mistero di sleepy hollow	tim burton	1999	horror	3.50
big fish	tim burton	2003	fantastico	3.10
la sposa cadavere	tim burton	2005	animazione	3.50
la fabbrica di cioccolato	tim burton	2005	fantastico	4.00
io non ho paura	gabriele salvatores	2003	drammatico	3.50
nirvana	gabriele salvatores	1997	fantascienza	3.00
mediterraneo	gabriele salvatores	1991	commedia	3.80
pulp fiction	quentin tarantino	1994	thriller	3.50
le iene	quentin tarantino	1992	thriller	4.00



Raggruppamento

- In un'interrogazione contenente una clausola GROUP BY, ogni tupla della relazione risultato rappresenta un gruppo di tuple della relazione su cui l'interrogazione è eseguita
- Nella query precedente i gruppi sono quattro: uno per ogni valore di regista
- Ad ognuno di questi gruppi sono applicate le funzioni di gruppo

Esempio



- Risultato:

regista	numF	numG	minV	avgV	maxV
emir kusturica	1	1	3.20	3.20	3.20
tim burton	8	5	3.00	3.58	4.00
gabriele salvatores	3	3	3.00	3.43	3.80
quentin tarantino	2	1	3.50	3.75	4.00

Esempio



- Il numero di figli di ciascun padre:

```
SELECT count(*) as NumFigli  
FROM Paternita  
GROUP BY padre;
```

Provate voi....

- Il numero di figli di ciascun padre:



Esempio



Film

titolo	regista	anno	genere	valutaz
underground	emir kusturica	1995	drammatico	3.20
edward mani di forbice	tim burton	1990	fantastico	3.60
nightmare before christmas	tim burton	1993	animazione	4.00
ed wood	tim burton	1994	drammatico	4.00
mars attacks	tim burton	1996	fantascienza	3.00
il mistero di sleepy hollow	tim burton	1999	horror	3.50
big fish	tim burton	2003	fantastico	3.10
la sposa cadavere	tim burton	2005	animazione	3.50
la fabbrica di cioccolato	tim burton	2005	fantastico	4.00
io non ho paura	gabriele salvatores	2003	drammatico	3.50
nirvana	gabriele salvatores	1997	fantascienza	3.00
mediterraneo	gabriele salvatores	1991	commedia	3.80
pulp fiction	quentin tarantino	1994	thriller	3.50
le iene	quentin tarantino	1992	thriller	4.00



Raggruppamento e proiezione

- Se partizioniamo i film per regista, non ha senso richiedere che venga restituito il titolo o la valutazione:

```
SELECT regista, titolo, valutaz
```

```
FROM Film
```

```
GROUP BY regista;
```

- Dopo aver eseguito il raggruppamento di Film su regista abbiamo quattro tuple:
 - ognuna corrisponde ad un valore di regista
 - i valori di titolo e valutaz corrispondenti sono tanti!

Raggruppamento e proiezione



- La clausola di proiezione di un'interrogazione contenente la clausola GROUP BY può solo includere:
 - una o più colonne tra le colonne che compaiono nella clausola GROUP BY
 - funzioni di gruppo



Maternità

Madre	Figlio
Luisa	Maria
Luisa	Luigi
Anna	Olga
Anna	Filippo
Maria	Andrea
Maria	Aldo

Paternità

Padre	Figlio
Sergio	Franco
Luigi	Olga
Luigi	Filippo
Franco	Andrea
Franco	Aldo

Persone

Nome	Età	Reddito
Andrea	27	21
Aldo	25	15
Maria	55	42
Anna	50	35
Filippo	26	30
Luigi	50	40
Franco	60	20
Olga	30	41
Sergio	85	35
Luisa	75	87

Raggruppamento



- Per ogni padre, il suo nome, il suo reddito e la media dei redditi dei figli



Raggruppamento

- Query scorretta:

```
SELECT Padre, AVG(f.Reddito), p.Reddito  
FROM Persone f JOIN Paternita on Figlio = Nome JOIN  
      Persone p ON Padre =p.Nome  
GROUP BY Padre;
```

- Query corretta:

```
SELECT Padre, AVG(f.Reddito), p.Reddito  
FROM Persone f JOIN Paternita on Figlio = Nome JOIN  
      Persone p ON Padre =p.Nome  
GROUP BY Padre, p.Reddito;
```




Raggruppamento

- Artificio sintattico
 - gli attributi univocamente determinati da attributi già presenti nella clausola GROUP BY possono essere aggiunti **senza alterare il risultato**



Raggruppamento

- Più colonne possono essere usate per definire gruppi
- Le funzioni di gruppo possono essere usate anche in presenza di join



Raggruppamento

- Per ogni regista che ha girato almeno due film prima del 2000, determinare quanti sono tali film, di quanti generi diversi e la valutazione minima, media e massima di tali film:
 - la condizione è definita su valori aggregati
 - non è possibile usare la clausola WHERE



Condizioni sui gruppi

- La clausola **HAVING** permette di specificare condizioni su valori aggregati
- **HAVING:**
 - e' seguita da un predicato o una combinazione Booleana di predicati
 - tali predicati possono essere solo predicati che coinvolgono funzioni di gruppo

Esempio



Film

titolo	regista	anno	genere	valutaz
underground	emir kusturica	1995	drammatico	3.20
edward mani di forbice	tim burton	1990	fantastico	3.60
nightmare before christmas	tim burton	1993	animazione	4.00
ed wood	tim burton	1994	drammatico	4.00
mars attacks	tim burton	1996	fantascienza	3.00
il mistero di sleepy hollow	tim burton	1999	horror	3.50
big fish	tim burton	2003	fantastico	3.10
la sposa cadavere	tim burton	2005	animazione	3.50
la fabbrica di cioccolato	tim burton	2005	fantastico	4.00
io non ho paura	gabriele salvatores	2003	drammatico	3.50
nirvana	gabriele salvatores	1997	fantascienza	3.00
mediterraneo	gabriele salvatores	1991	commedia	3.80
pulp fiction	quentin tarantino	1994	thriller	3.50
le iene	quentin tarantino	1992	thriller	4.00



Esempio

Per ogni regista che ha girato almeno due film prima del 2000, determinare quanti sono tali film, di quanti generi diversi e la valutazione minima, media e massima di tali film:

```
SELECT regista, COUNT(*) AS numF,  
        COUNT(DISTINCT genere) AS numG,  
        MIN(valutaz) AS minV, AVG(valutaz) AS avgV,  
        MAX(valutaz) AS maxV  
FROM Film  
WHERE anno < 2000  
GROUP BY regista  
HAVING COUNT(*) >= 2;
```

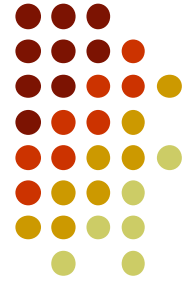
Esempio



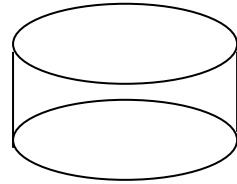
Risultato:

regista	numF	numG	minV	avgV	maxV
tim burton	5	5	3.00	3.62	4.00
gabriele salvatores	2	2	3.00	3.40	3.80
quentin tarantino	2	1	3.50	3.75	4.00

Raggruppamento

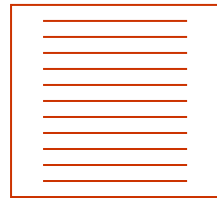


base di dati



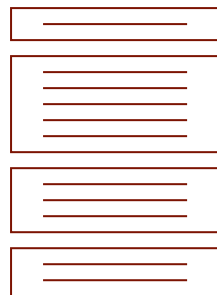
1. ↓

film girati
prima del 2000



2. ↓

film girati
prima del 2000
raggruppati
per regista



3. →



risultato

4. ↑



film girati prima del 2000
raggruppati per regista,
considerando solo registi
che ne hanno girato
almeno due

Maternita

Madre	Figlio
Luisa	Maria
Luisa	Luigi
Anna	Olga
Anna	Filippo
Maria	Andrea
Maria	Aldo

Paternita

Padre	Figlio
Sergio	Franco
Luigi	Olga
Luigi	Filippo
Franco	Andrea
Franco	Aldo

Persone

Nome	Età	Reddito
Andrea	27	21
Aldo	25	15
Maria	55	42
Anna	50	35
Filippo	26	30
Luigi	50	40
Franco	60	20
Olga	30	41
Sergio	85	35
Luisa	75	87



WHERE o HAVING?



I padri i cui figli sotto i 30 anni hanno un reddito medio maggiore di 25:

```
SELECT Padre  
FROM Persone JOIN Paternita ON Figlio = Nome  
WHERE Eta < 30  
GROUP BY Padre  
HAVING AVG(Reddito) > 25;
```

Esempio



I padri che hanno almeno due figli:

```
SELECT Padre  
FROM Paternita  
GROUP BY Padre  
HAVING COUNT(*) >= 2;
```

Alternativa senza COUNT



I padri che hanno almeno due figli:

```
SELECT Padre  
FROM Paternita p JOIN Paternita x ON (p.Padre = x.Padre)  
WHERE p.Figlio <> x.Figlio;
```