

Universita' di Milano Bicocca  
Corso di Basi di dati 1 in eLearning  
C. Batini  
7. SQL DML  
7.7 Operatore di Raggruppamento

# Raggruppamenti: l'operatore GROUP BY

# Limiti degli operatori aggregati

- Fino a questo momento le funzioni di aggregazione (COUNT, ecc.) si applicano all' intero risultato della interrogazione.

Le funzioni di aggregazione possono essere applicate anche a partizioni delle relazioni, cioè a gruppi di tuple.

# Operatori aggregati e raggruppamenti

Le funzioni di aggregazione possono essere applicate a partizioni delle relazioni, cioè a gruppi di tuple. Si usa in questo caso l'operatore **GROUP BY**, con sintassi:

**GROUP BY** listaAttributi

che viene inserito alla fine della **SELECT**

```
SELECT target list  
FROM ...  
WHERE ...  
GROUP BY listaAttributi
```

# Semantica delle interrogazioni con operatori aggregati e raggruppamenti

1. Esegui la interrogazione senza tener conto la GROUP BY e gli operatori aggregati
2. Raggruppa le righe che hanno stessi valori per gli attributi che compaiono nella listaAttributi della GROUP BY
3. Applica l'operatore aggregato a ciascun gruppo di n-ple

# Esempio utilizzato nel seguito

Sia data la relazione

Paternalita' (Padre, Figlio, Reddito)

dove Reddito e' il reddito del figlio

# Operatori aggregati e raggruppamenti

- Trovare il numero di figli di ciascun padre che hanno reddito maggiore di 25

```
select Padre, count(*) AS NumFigli  
from Paternita  
where Reddito > 25  
group by Padre
```

Padre	Figlio	Reddito
Sergio	Franco	50
Luigi	Olga	30
Luigi	Filippo	20
Franco	Andrea	40
Franco	Aldo	40



1. Esegui la interrogazione "in giallo"  
senza GROUP BY e senza operatori aggregati

1. select Padre, count(\*) AS NumFigli  
from Paternita  
Where Reddito > 25  
group by Padre

Domanda 7.7.1

Quale e' il risultato?

Padre	Figlio	Reddito
Sergio	Franco	50
Luigi	Olga	30
Luigi	Filippo	20
Franco	Andrea	40
Franco	Aldo	40



Padre	Figlio	Reddito
Sergio	Franco	50
Luigi	Olga	30
Franco	Andrea	40
Franco	Aldo	40



## 2. Raggruppa le righe che hanno stessi valori per gli attributi nella GROUP BY

```
select Padre, count(*) AS  
NumFigli  
from Paternita  
Where Reddito > 25  
2. group by Padre
```

Domanda 7.7.2

Quale e' il risultato?

Padre	Figlio	Reddito
Sergio	Franco	50
Luigi	Olga	30
Franco	Andrea	40
Franco	Aldo	40



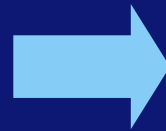
Padre	Figlio
Sergio	Franco
Luigi	Olga
Franco	Andrea
Franco	Aldo

### 3. Applica l'operatore aggregato a ciascun gruppo

```
select Padre, 3. count(*) AS NumFigli  
from Paternita  
Where Reddito > 25  
group by Padre
```

Domanda 7.7.3  
Quale e' il risultato?

Padre	Figlio
Sergio	Franco
Luigi	Olga
Franco	Andrea
Franco	Aldo




Padre	NumFigli
Sergio	1
Luigi	1
Franco	2

# Regola da rispettare nella clausola GROUP BY

In una interrogazione che fa uso della GROUP BY puo' comparire come argomento della SELECT, e quindi come insieme di attributi nella target list, solamente un sottoinsieme degli attributi che compaiono nella clausola GROUP BY

SELECT target list  
FROM ...  
WHERE ...  
GROUP BY insieme di attributi di aggregazione



Vediamo perche'

# Operatori aggregati e raggruppamenti

## Esempio 1

```
select Padre, count(*) AS NumFigli  
from Paternita  
Where Reddito > 50  
group by Padre
```

Domanda 7.7.4 Esprimi in linguaggio naturale  
cosa calcola questa interrogazione.

Risposta

Trova il numero di figli di ciascun padre con  
reddito maggiore di 50

Va bene

## Esempio 2

Consideriamo questa relazione e sua istanza  
Esaminala con attenzione

Cognome	Dipartimento	Ufficio
Rossi	Amministrazione	10
Verdi	Produzione	20
Neri	Direzione	30
Gini	Amministrazione	15
Dini	Amministrazione	25
Peri	Produzione	40
Turi	Direzione	50

## Esempio 2

Nello schema

Impiegato (Cognome, Dipartimento, Ufficio)

L'interrogazione

```
Select Dipartimento, Ufficio, Count (Ufficio)
```

```
From Impiegato
```

```
Group by Dipartimento
```

Non e' corretta (e, nota, non rispetta la regola!)

Infatti applicando la group by e poi la select si  
ottiene ...

Select Dipartimento, Ufficio, Count (Ufficio)  
From Impiegato  
Group by Dipartimento

## Situazione di partenza

Cognome	Dipartimento	Ufficio
Rossi	Amministrazione	10
Verdi	Produzione	20
Neri	Direzione	30
Gini	Amministrazione	15
Dini	Amministrazione	25
Peri	Produzione	40
Turi	Direzione	50

Select Dipartimento, Ufficio, Count (Ufficio)  
From Impiegato  
Group by Dipartimento

Raggruppamenti  
generati - 1

Cognome	Dipartimento	Ufficio
Rossi	Amministrazione	10
Verdi	Produzione	20
Neri	Direzione	30
Gini	Amministrazione	15
Dini	Amministrazione	25
Peri	Produzione	40
Turi	Direzione	50



Select Dipartimento, Ufficio, Count (Ufficio)

From Impiegato

Group by Dipartimento

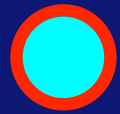
Raggruppamenti  
generati - 2

Cognome	Dipartimento	Ufficio
Rossi	Amministrazione	10
Gini	Amministrazione	15
Dini	Amministrazione	25
Verdi	Produzione	20
Peri	Produzione	40
Neri	Direzione	30
Turi	Direzione	50

# Risultato finale

Dipartimento	Ufficio	Numero uffici
Amministrazione	10	3
	15	
	25	
Produzione	20	2
	40	
Direzione	30	2
	50	

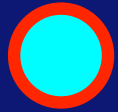
Che non e' coerente con la definizione di istanza nel modello relazionale, in cui nelle n-ple non possono comparire multivalori



## Altro esempio - 3

Se lo scopo della interrogazione era trovare per ogni dipartimento il numero di uffici, va bene invece l'interrogazione

```
Select Dipartimento, count (Ufficio)
From Impiegato
Group by Dipartimento
```



## La restrizione e' eccessiva ...

Consideriamo la interrogazione

```
Select Dipartimento, count (*), Citta'  
From Impiegato Join Dipartimento  
on (Dipartimento = Nome)
```

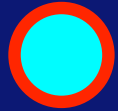
Group by Dipartimento

Eseguita sullo schema

Dipartimento (Nome, Citta')

Impiegato (Cognome, Dipartimento)

Restituirebbe, poiche' c'e' una sola citta' per dipartimento, i dipartimenti, il numero di impiegati dei dipartimenti, e la citta' del dipartimento.

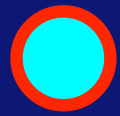


Ma ....

- Non e' accettata perche' la implementazione sarebbe piu' complessa, dovendo il compilatore verificare quali sono gli attributi ammissibili

D'altra parte, si puo' sempre esprimere con un insieme di attributi ridondante

```
Select Dipartimento, count (*), Citta'  
From Impiegato Join Dipartimento  
on (Dipartimento = Nome)  
Group by Dipartimento, Citta'
```



## Raggruppamenti e target list: altro esempio

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
```

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
```

# Concetti introdotti

- Operatore *GROUP BY* di raggruppamento