

# Basi di dati cod. 861II [9 CFU]

Corso di Laurea in Ingegneria Informatica

---

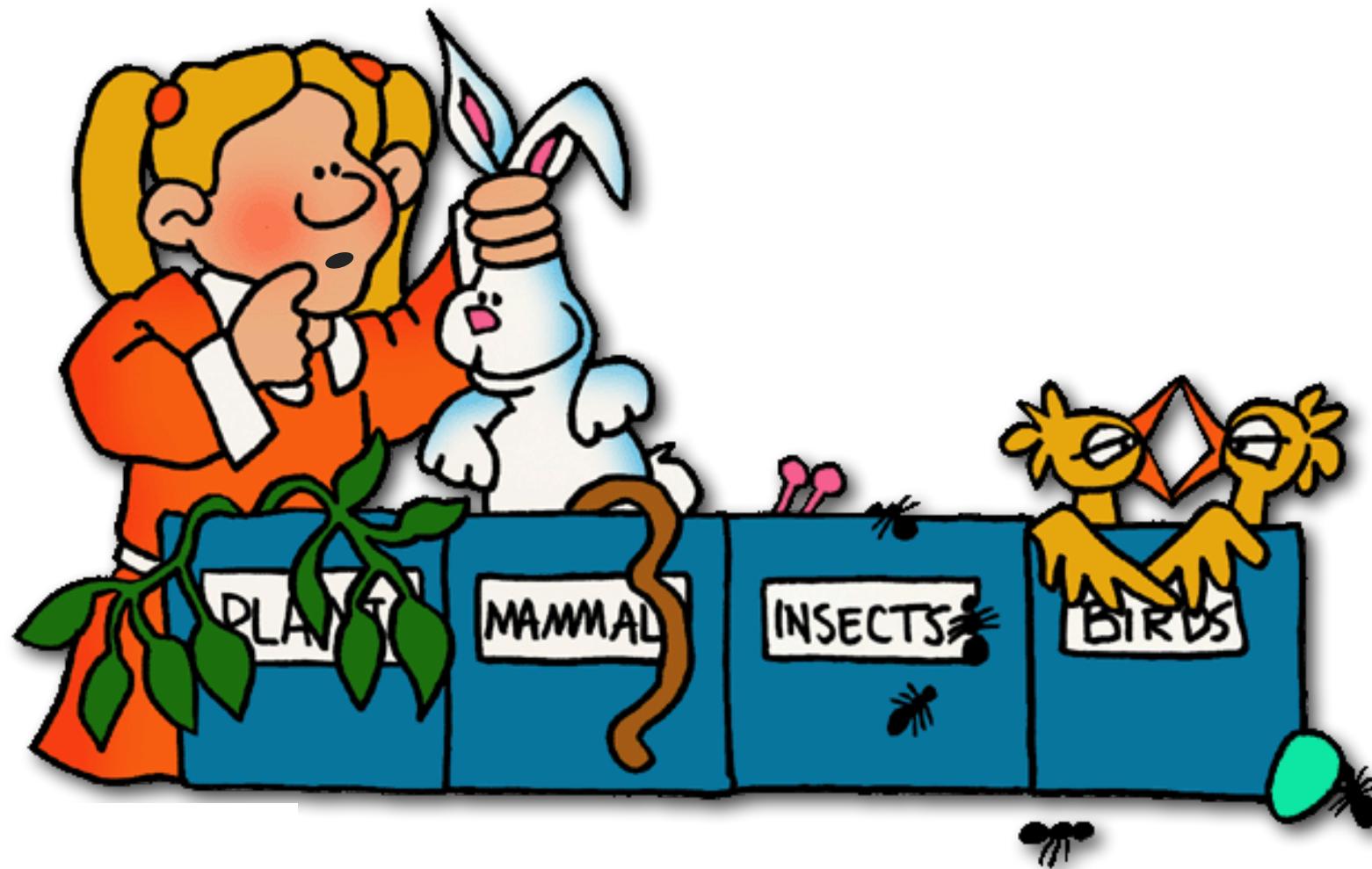
**Oracle MySQL**  
A.A. 2022-2023

Francesco Pistolesi  
Dipartimento di Ingegneria dell'Informazione  
Università di Pisa  
[francesco.pistolesi@unipi.it](mailto:francesco.pistolesi@unipi.it)



# Raggruppamento

---



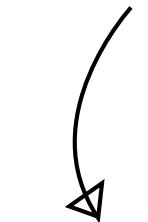
# Raggruppamento

---

la tabella target viene frammentata

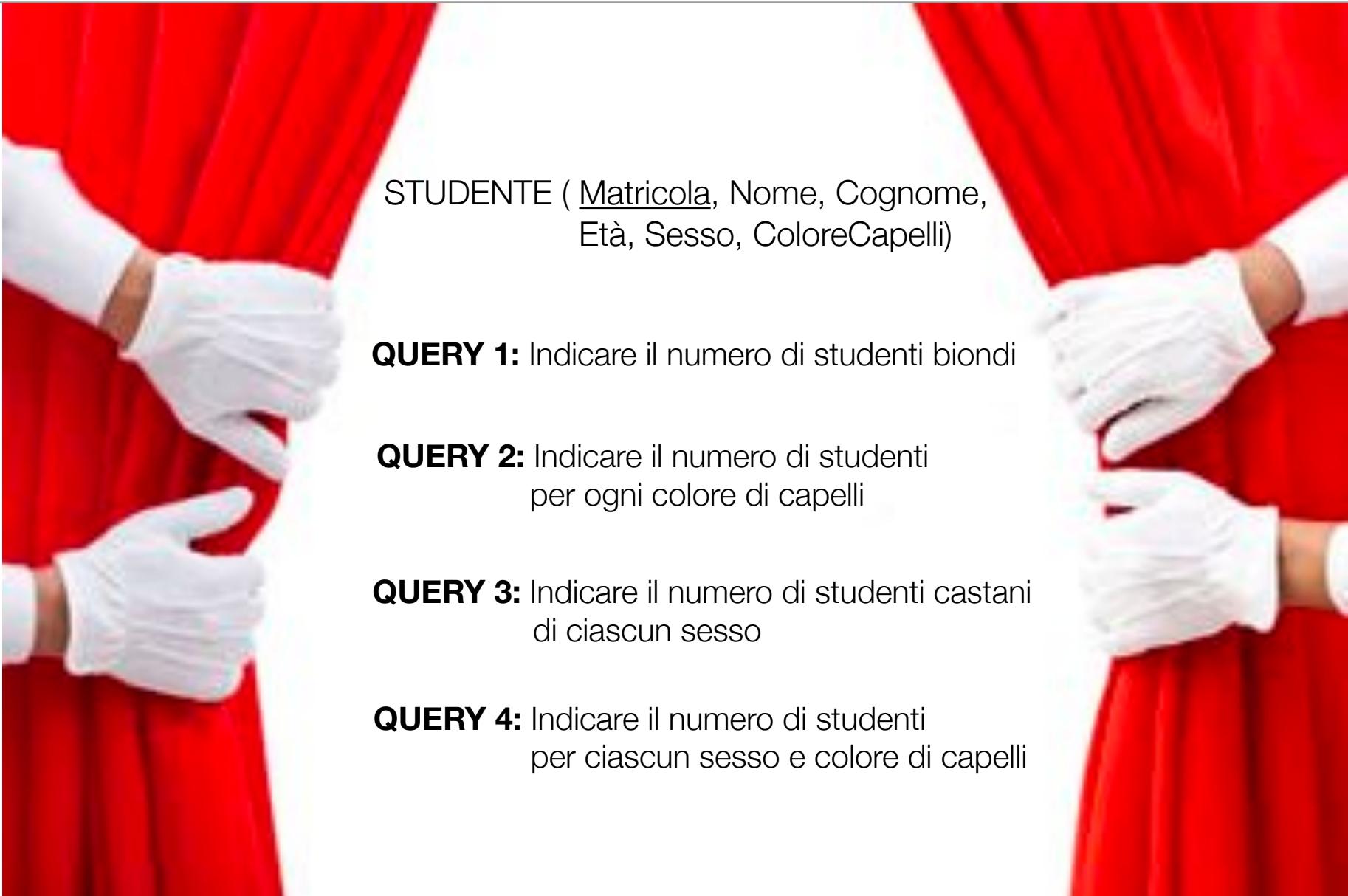


Suddivide un insieme di record in **gruppi di record**; in ogni gruppo, il valore di un attributo **è costante in tutti i record**



anche più di un attributo

# Siparietto



STUDENTE ( Matricola, Nome, Cognome,  
Età, Sesso, ColoreCapelli)

**QUERY 1:** Indicare il numero di studenti biondi

**QUERY 2:** Indicare il numero di studenti  
per ogni colore di capelli

**QUERY 3:** Indicare il numero di studenti castani  
di ciascun sesso

**QUERY 4:** Indicare il numero di studenti  
per ciascun sesso e colore di capelli

# Raggruppamento: un semplice esempio

---

Data **ogni specializzazione**, indicarne il nome e la parcella media dei medici

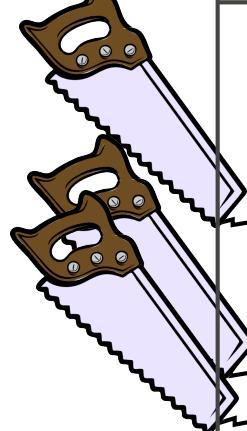
**MEDICO**

<b>Matricola</b>	<b>Cognome</b>	<b>Nome</b>	<b>Specializzazione</b>	<b>Parcella</b>
014	Indachi	Loredana	Cardiologia	180
003	Gialli	Franco	Otorinolaringoiatria	120
010	Celesti	Clelia	Neurologia	200
006	Bianchi	Ada	Ortopedia	160
002	Verdi	Luigi	Otorinolaringoiatria	150
017	Amaranti	Adevane	Cardiologia	250
001	Rossi	Mario	Otorinolaringoiatria	180
007	Rosi	Alvaro	Ortopedia	180
015	Ciani	Gualtiero	Cardiologia	230
018	Terra di Siena	Bruciata	Cardiologia	260

# Raggruppamento: un semplice esempio

Data **ogni specializzazione**, indicarne il nome e la parcella media dei medici

**MEDICO**



<b>Matricola</b>	<b>Cognome</b>	<b>Nome</b>	<b>Specializzazione</b>	<b>Parcella</b>
014	Indachi	Loredana	Cardiologia	180
003	Amaranti	Adevane	Cardiologia	250
010	Ciani	Gualtiero	Cardiologia	230
006	Terra di Siena	Bruciata	Cardiologia	260
002	Gialli	Franco	Otorinolaringoiatria	120
017	Verdi	Luigi	Otorinolaringoiatria	150
001	Rossi	Mario	Otorinolaringoiatria	180
012	Celesti	Clelia	Neurologia	200
015	Bianchi	Ada	Ortopedia	160
018	Rosi	Alvaro	Ortopedia	180

# Raggruppamento: un semplice esempio

---

Data **ogni specializzazione**, indicarne il nome e la parcella media dei medici

Matricola	Cognome	Nome	Specializzazione	Parcella
014	Indachi	Loredana	Cardiologia	180
003	Amaranti	Adevane	Cardiologia	250
010	Ciani	Gualtiero	Cardiologia	230
006	Terra di Siena	Bruciata	Cardiologia	260

Matricola	Cognome	Nome	Specializzazione	Parcella
002	Gialli	Franco	Otorinolaringoiatria	120
017	Verdi	Luigi	Otorinolaringoiatria	150
001	Rossi	Mario	Otorinolaringoiatria	180

Matricola	Cognome	Nome	Specializzazione	Parcella
12	Celesti	Clelia	Neurologia	200

Matricola	Cognome	Nome	Specializzazione	Parcella
015	Bianchi	Ada	Ortopedia	160
018	Rosi	Alvaro	Ortopedia	180

# Raggruppamento: un semplice esempio

Matricola	Cognome	Nome	Specializzazione	Parcella
014	Indachi	Loredana	Cardiologia	180
003	Amaranti	Adevane	Cardiologia	250
010	Ciani	Gualtiero	Cardiologia	230
006	Terra di Siena	Bruciata	Cardiologia	260

$$\text{AVG}(\text{Parcella}) = 230$$

Matricola	Cognome	Nome	Specializzazione	Parcella
002	Gialli	Franco	Otorinolaringoiatria	120
017	Verdi	Luigi	Otorinolaringoiatria	150
001	Rossi	Mario	Otorinolaringoiatria	180

$$\text{AVG}(\text{Parcella}) = 150$$

# Raggruppamento: un semplice esempio

Matricola	Cognome	Nome	Specializzazione	Parcella
12	Celesti	Clelia	Neurologia	200

**AVG(Parcella) = 200**



Matricola	Cognome	Nome	Specializzazione	Parcella
015	Bianchi	Ada	Ortopedia	160
018	Rosi	Alvaro	Ortopedia	180

**AVG(Parcella) = 170**



# Raggruppamento: un semplice esempio

---

Data **ogni specializzazione**, indicarne il nome e la parcella media dei medici

Specializzazione	ParcellaMedia
Cardiologia	230
Otorinolaringoiatria	150
Neurologia	200
Ortopedia	170



ogni gruppo collassa in un unico record, sempre!



OK, adesso basta con i  
disegnini...come cavolo si fa questa  
roba in MySQL?????

# Raggruppamento in MySQL

Indicare la parcella media dei medici **di ciascuna specializzazione**

```
SELECT Specializzazione, AVG(Parcella) AS ParcellaMedia  
FROM Medico  
GROUP BY Specializzazione;
```

assume lo stesso valore  
in un gruppo

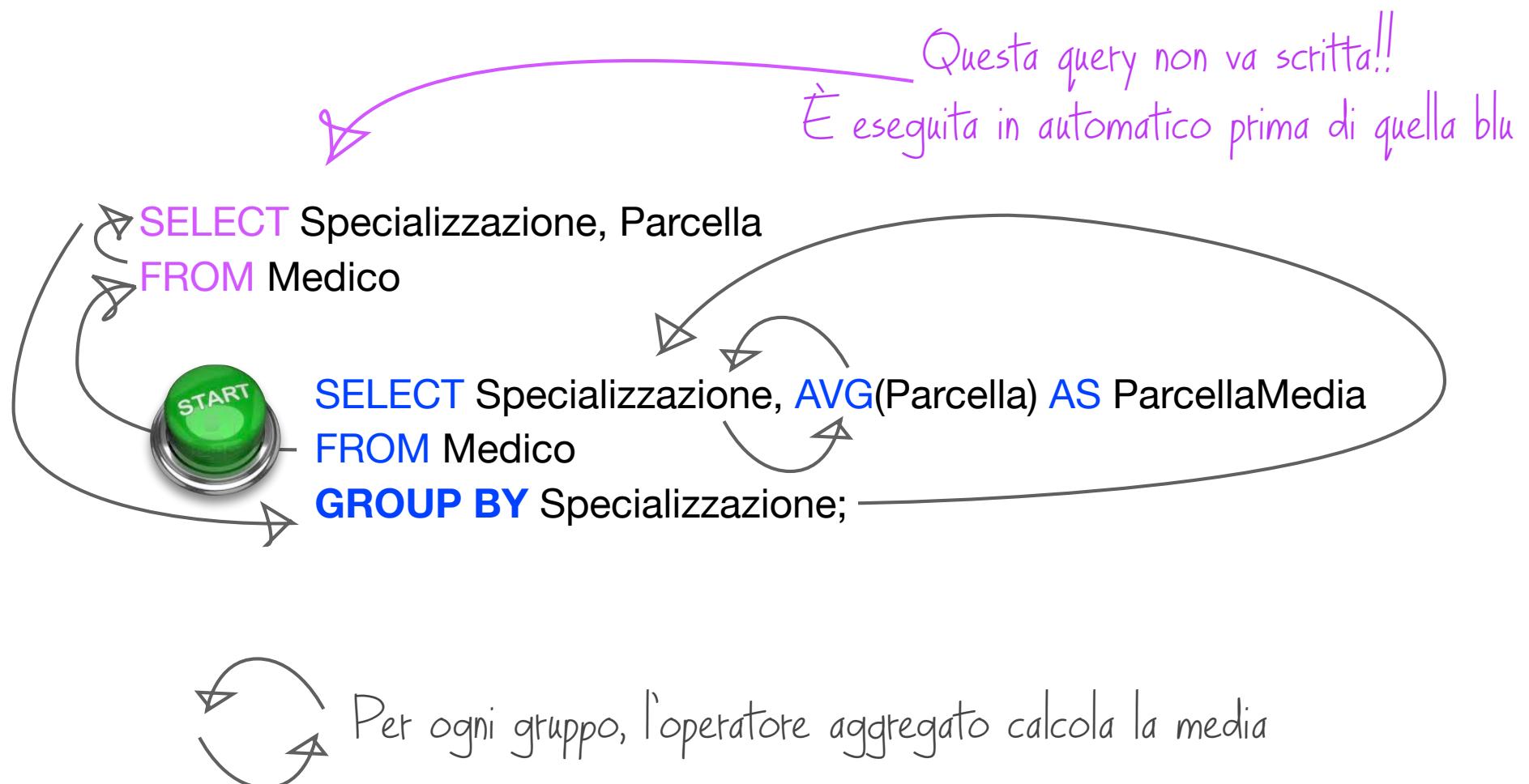
l'operatore è applicato gruppo per gruppo  
(calcola un valore riassuntivo per il gruppo!)

assume valori diversi in un gruppo,  
non si può proiettare senza aggregazione

la clausola "group by" permette di frammentare la tabella

# Raggruppamento in MySQL: processazione

Indicare la parcella media dei medici **di ciascuna specializzazione**



# Raggruppamento: regola fondamentale

---



# Raggruppamento e proiezione

---

Per ogni specializzazione, calcolare la parcella media dei suoi medici

```
SELECT Specializzazione, AVG(Parcella) AS ParcellaMedia  
FROM Medico  
GROUP BY Specializzazione;
```

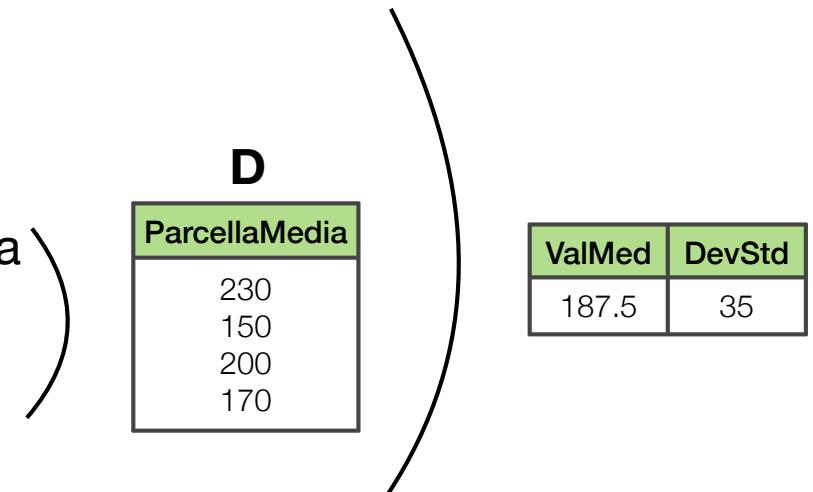
Specializzazione	ParcellaMedia
Cardiologia	230
Otorinolaringoiatria	150
Neurologia	200
Ortopedia	170

L'attributo di raggruppamento nella proiezione permette di capire a quale specializzazione si riferisce la parcella media

# Si può non proiettare l'attributo di raggruppamento?

Restituire valor medio e deviazione standard delle parcella medie  
dei medici delle varie specializzazioni

```
SELECT AVG(D.ParcellaMedia) AS ValMed,  
       STDDEV(D.ParcellaMedia) AS DevStd,  
FROM(  
    SELECT AVG(Parcella) AS ParcellaMedia  
    FROM Medico  
    GROUP BY Specializzazione  
) AS D;
```



In questo caso, si può omettere l'attributo di raggruppamento nella derived table:  
non serve risalire alla specializzazione a cui una parcella media appartiene!

# Il male è sempre in agguato...

---

Per ogni specializzazione medica, indicarne il nome, la parcella minima  
**e il cognome del medico a cui appartiene**



# Guai in vista



Per ogni specializzazione medica, indicarne il nome, la parcella minima  
**e il cognome del medico a cui appartiene**,

~~SELECT Specializzazione, MIN(Parcella), Cognome  
FROM Medico  
GROUP BY Specializzazione;~~

Non è connesso al medico  
avente patcella minima...  
(e non è costante in un gruppo!)

# Cosa significa???

Per ogni specializzazione medica, indicarne il nome, la parcella minima  
**e il cognome del medico a cui appartiene**

```
SELECT Specializzazione, MIN(Parcella), Cognome  
FROM Medico  
GROUP BY Specializzazione;
```

Esempio: il gruppo dei cardiologi

Matricola	Cognome	Nome	Specializzazione	Parcella
014	Indachi	Loredana	Cardiologia	180
003	Amaranti	Adevane	Cardiologia	250
010	Ciani	Gualtiero	Cardiologia	230
006	Terra di Siena	Bruciata	Cardiologia	260

Specializzazione	MIN(Parcella)
Cardiologia	180



Cognome??

GROUP BY → MIN(Parcella)

# Per non dimenticare...

---

La proiezione può contenere  
solo attributi presenti

nel predicato di raggruppamento

(perché sono i soli ad assumere  
un valore costante in ogni gruppo).

Gli altri attributi proiettati  
devono essere argomento di  
funzioni di aggregazione!!!

# Sconfiggiamo Malefica!

---



# Sconfiggiamo Malefica!



`SELECT Specializzazione, MIN(Parcella)  
FROM Medico  
GROUP BY Specializzazione`



???

Per ogni specializzazione medica, indicarne il nome, la parcella minima  
**e il cognome del medico a cui appartiene**

`SELECT Cognome  
FROM Medico  
WHERE Parcella = "parcella minima della sua specializzazione"`



# Sconfiggiamo Malefica!



```
SELECT Specializzazione, MIN(Parcella)
      AS ParcellaMin
   FROM Medico
GROUP BY Specializzazione
```

Specializzazione	ParcellaMin
Cardiologia	180
Otorinolaringoiatria	100
Neurologia	200
Ortopedia	160

```
SELECT Cognome, Specializzazione, Parcella
   FROM Medico
```

Cognome	Specializzazione	Parcella
Indachi	Cardiologia	180
Gialli	Otorinolaringoiatria	200
Celesti	Neurologia	200
Bianchi	Ortopedia	160
Verdi	Otorinolaringoiatria	150
Amaranti	Cardiologia	250
Rossi	Otorinolaringoiatria	100
Rosi	Ortopedia	180

# Sconfiggiamo Malefica!



Per ogni specializzazione medica, indicarne il nome, la parcella minima  
**e il cognome del medico a cui appartiene**

Cognome	Specializzazione	Parcella
Indachi	Cardiologia	180
Gialli	Otorinolaringoiatria	200
Celesti	Neurologia	200
Bianchi	Ortopedia	160
Verdi	Otorinolaringoiatria	150
Amaranti	Cardiologia	250
Rossi	Otorinolaringoiatria	100
Rosi	Ortopedia	180



Specializzazione	Parcella Min
Cardiologia	180
Otorinolaringoiatria	100
Neurologia	200
Ortopedia	160

Join naturale

# Sconfiggiamo Malefica!



Per ogni specializzazione medica, indicarne il nome, la parcella minima  
**e il cognome del medico a cui appartiene**

Cognome	Specializzazione	Parcella	ParcellaMin
Indachi	Cardiologia	180	180
Gialli	Otorinolaringoiatria	200	100
Celesti	Neurologia	200	200
Bianchi	Ortopedia	160	160
Verdi	Otorinolaringoiatria	150	100
Amaranti	Cardiologia	250	180
Rossi	Otorinolaringoiatria	100	100
Rosi	Ortopedia	180	160

# Sconfiggiamo Malefica!



Per ogni specializzazione medica, indicarne il nome, la parcella minima  
**e il cognome del medico a cui appartiene**

```
SELECT M.Specializzazione,  
       D.ParcellaMinima,  
       M.Cognome  
FROM Medico M  
      NATURAL JOIN  
(  
    SELECT Specializzazione, MIN(Parcella) AS ParcellaMinima  
    FROM Medico  
    GROUP BY Specializzazione  
) AS D  
WHERE M.Parcella = D.ParcellaMinima;
```

# Condizioni sui gruppi

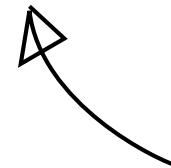
---



# Condizioni sui gruppi

---

perché esprimono una caratteristica di un gruppo, cioè globale, sommativa, riepilogativa degli attributi variabili (non di raggruppamento) dei record che lo compongono



Sono espresse esclusivamente tramite **operatori di aggregazione** e permettono di **scartare gruppi**, qualora non siano soddisfatte



sono controllate grupper per grupper,  
non record per record

# Condizioni sui gruppi: esempio

---

Indicare le specializzazioni della clinica **con più di due medici**



è evidente che si deve ragionare sul numero  
di medici di ciascuna specializzazione

# Condizioni sui gruppi: esempio

---

Indicare le specializzazioni della clinica **con più di due medici**

MEDICO

Cognome	Specializzazione	Città
Indachi	Cardiologia	Pisa
Gialli	Otorinolaringoiatria	Pisa
Celesti	Neurologia	Firenze
Bianchi	Ortopedia	Milano
Verdi	Otorinolaringoiatria	Pisa
Amaranti	Cardiologia	Pisa
Rossi	Otorinolaringoiatria	Pisa
Rosi	Ortopedia	Roma
Terra di Siena	Cardiologia	Siena

# Condizioni sui gruppi

Indicare le specializzazioni della clinica **con più di due medici**

Rossi	Otorinolaringoiatria	Pisa
Gialli	Otorinolaringoiatria	Pisa
Verdi	Otorinolaringoiatria	Pisa



COUNT(\*) = 3

Bianchi	Ortopedia	Milano
Rosi	Ortopedia	Roma



COUNT(\*) = 2

Amaranti	Cardiologia	Pisa
Indachi	Cardiologia	Pisa
Tetra di Siena	Cardiologia	Siena



COUNT(\*) = 3

Celesti	Neurologia	Firenze
---------	------------	---------



COUNT(\*) = 1

# Condizioni sui gruppi

Indicare le specializzazioni della clinica **con più di due medici**

Rossi	Otorinolaringoiatria	Pisa
Gialli	Otorinolaringoiatria	Pisa
Verdi	Otorinolaringoiatria	Pisa

COUNT(\*) = 3

Bianchi	Chirurgia	Milano
Zanini	Cirurgia	Roma

COUNT(\*) = 2

Amaranti	Cardiologia	Pisa
Indachi	Cardiologia	Pisa
Tetra di Siena	Cardiologia	Siena

COUNT(\*) = 3

Levi	Neurologia	Firenze
Castiglioni	Neurologia	Firenze

COUNT(\*) = 1

# Condizioni sui gruppi in MySQL

---

Indicare le specializzazioni della clinica **con più di due medici**

```
SELECT Specializzazione  
FROM Medico  
GROUP BY Specializzazione  
HAVING COUNT(*) > 2;
```

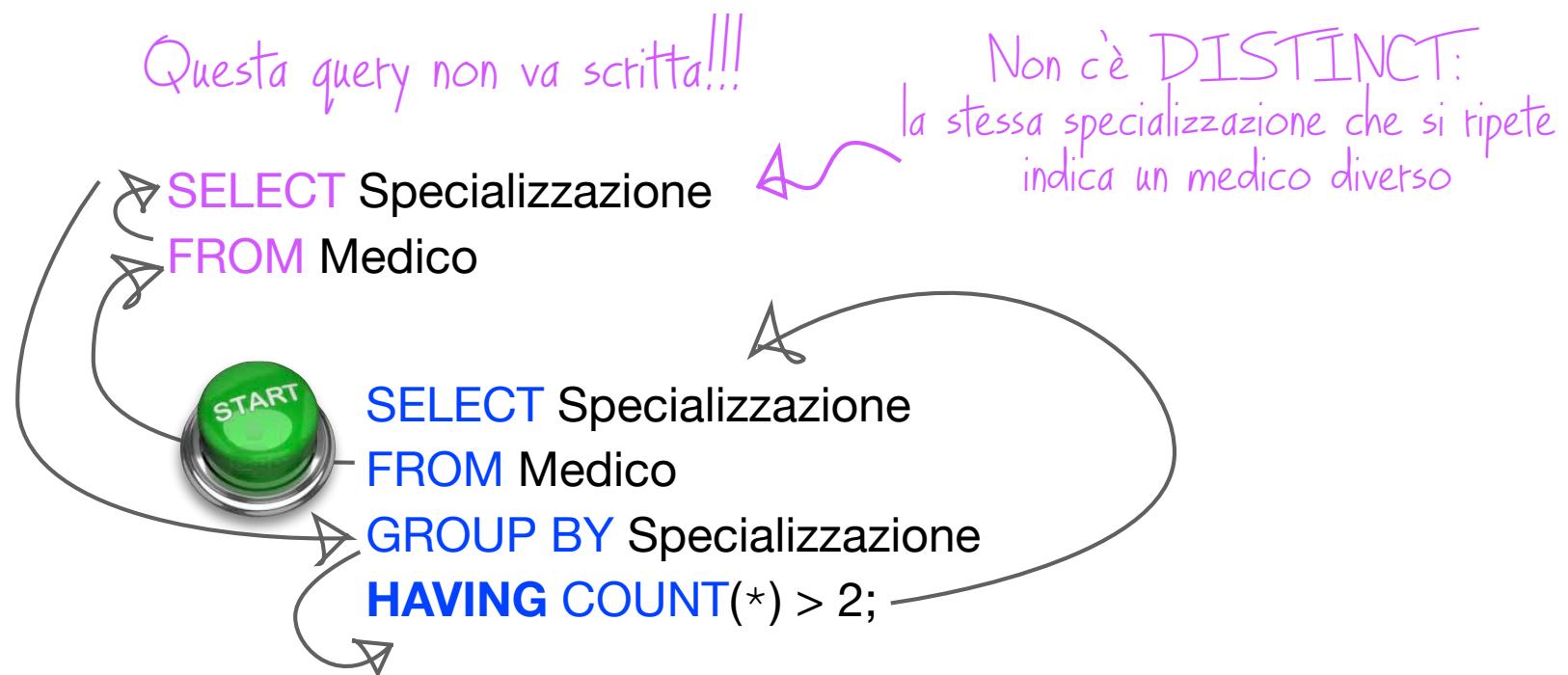
*ma serve distinct?/*



NO! Specializzazione è l'attributo di raggruppamento. È costante in ciascun gruppo. Che senso ha parlare di duplicati?

# Condizioni sui gruppi in SQL: processazione

Indicare le specializzazioni della clinica **con più di due medici**



# Si sono alleate contro di noi...

---

Indicare le specializzazioni con la più alta parcella media



# Errore blu

---

Indicare le specializzazioni con la più alta parcella media

```
SELECT Specializzazione  
FROM Medie  
GROUP BY Specializzazione  
HAVING MAX(AVG(Parcella));
```

Questo è il delirio!!!

Gli operatori di aggregazione non si annidano mai!  
La clausola having contiene sempre 2 parti connesse da  
un operatore di confronto (=, >, <, >=, <=)



# Ragioniamo con calma, cosa ci occorre?

---



Indicare le specializzazioni con la più alta parcella media

1. Media delle parcelle dei medici di ciascuna specializzazione
2. Calcolo della più alta parcella media
3. Specializzazioni con media parcelle uguale alla più alta parcella media



# Risoluzione: primo passo

Indicare le specializzazioni con la più alta parcella media

1. Media delle parcelle dei medici di ciascuna specializzazione

```
SELECT Specializzazione, AVG(Parcella)  
FROM Medico  
GROUP BY Specializzazione;
```



# Risoluzione: secondo passo

Indicare le specializzazioni con la più alta parcella media

2. Calcolo della più alta parcella media

```
SELECT MAX(D.MediaParcelle)
FROM
(
    SELECT M2.Specializzazione,
           AVG(M2.Parcella) AS MediaParcelle
    FROM Medico M2
    GROUP BY M2.Specializzazione
)
AS D;
```



# Soluzione

Indicare le specializzazioni con la più alta parcella media

```
SELECT M.Specializzazione ← Proietta tutti i dati merito  
FROM Medico M  
GROUP BY M.Specializzazione  
HAVING AVG(M.Parcella) =  
(  
    SELECT MAX(D.MediaParcelle)  
    FROM  
    (  
        SELECT M2.Specializzazione,  
              AVG(M2.Parcella) AS MediaParcelle  
        FROM Medico M2  
        GROUP BY M2.Specializzazione  
    )  
    AS D  
);
```

The code is annotated with handwritten numbers 1 and 2:

- Annotation 1 is located next to the innermost SELECT statement within the subquery, pointing to the part `AVG(M2.Parcella) AS MediaParcelle`.
- Annotation 2 is located next to the outer SELECT statement, pointing to the part `MAX(D.MediaParcelle)`.

# Condizioni sui gruppi vs. condizioni sui record

---

H  
A  
V  
I  
N  
G

C



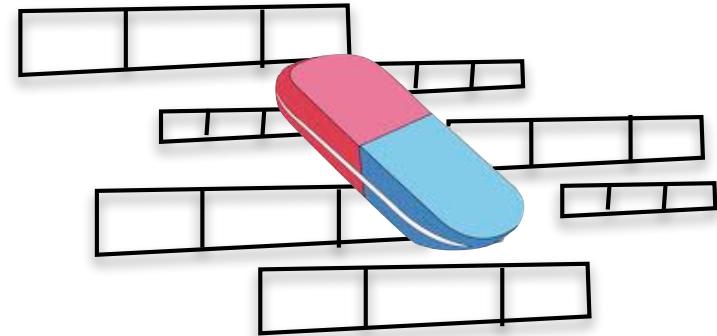
W  
H  
E  
R  
E  
T

E

# Condizioni sui gruppi vs. condizioni sui record

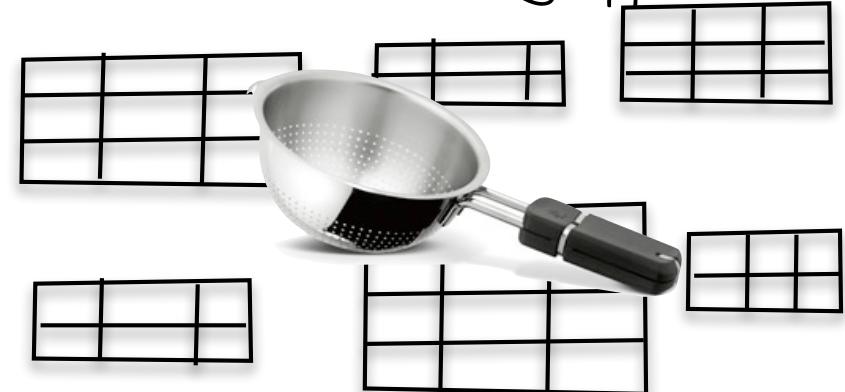
le condizioni nel **WHERE** sono applicate  
ai record **prima** del raggruppamento

eliminano record



le condizioni nell'**HAVING** sono applicate  
ai gruppi, **dopo** il raggruppamento

eliminano gruppi



# Condizioni sui gruppi vs. condizioni sui record

---



# Condizioni sui gruppi: regola



# Condizioni sui gruppi e sui record: esempio



# Condizioni sui gruppi e sui record: esempio

Indicare le specializzazioni con **più di due medici di Pisa**.

MEDICO

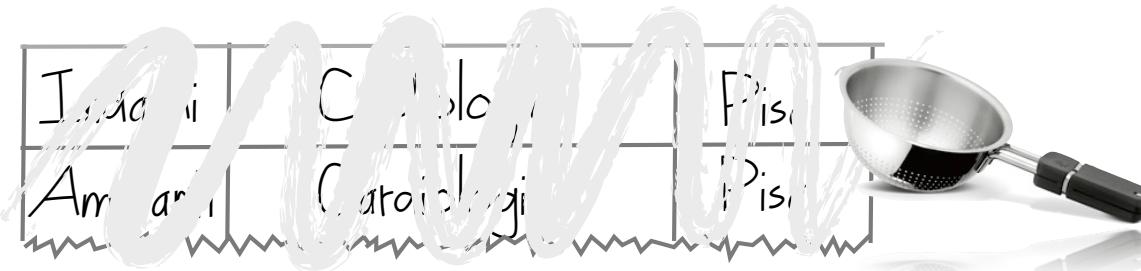
Cognome	Specializzazione	Città
Indachi	Cardiologia	Pisa
Gialli	Otorinolaringoiatria	Pisa
Verdi	Otorinolaringoiatria	Pisa
Amaranti	Cardiologia	Pisa
Rossi	Otorinolaringoiatria	Pisa
Tutti	Cardiologia	Cina



# Condizioni sui gruppi e sui record: esempio

Indicare le specializzazioni con **più di due medici di Pisa**.

Iacovi	Circolo	Pisa
Amaro	Circolo	Pisa



COUNT(\*) = 2

Rossi	Otorinolaringoiatria	Pisa
Gialli	Otorinolaringoiatria	Pisa
Verdi	Otorinolaringoiatria	Pisa

COUNT(\*) = 3

# Soluzione

---

Indicare le specializzazioni con **più di due medici di Pisa**.

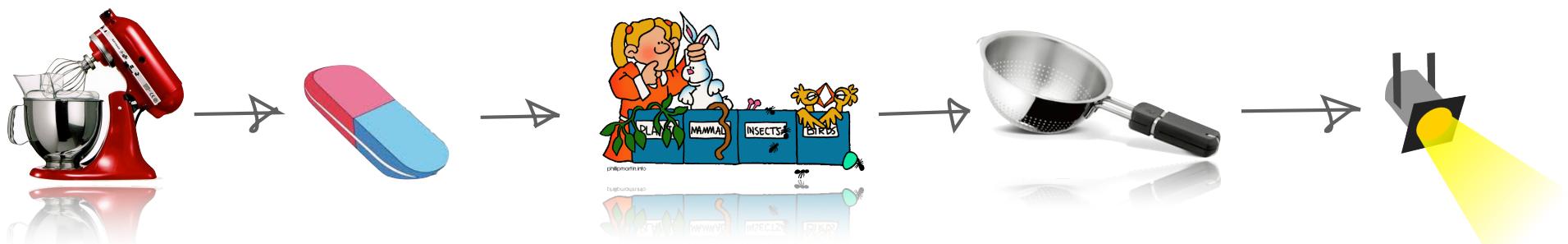
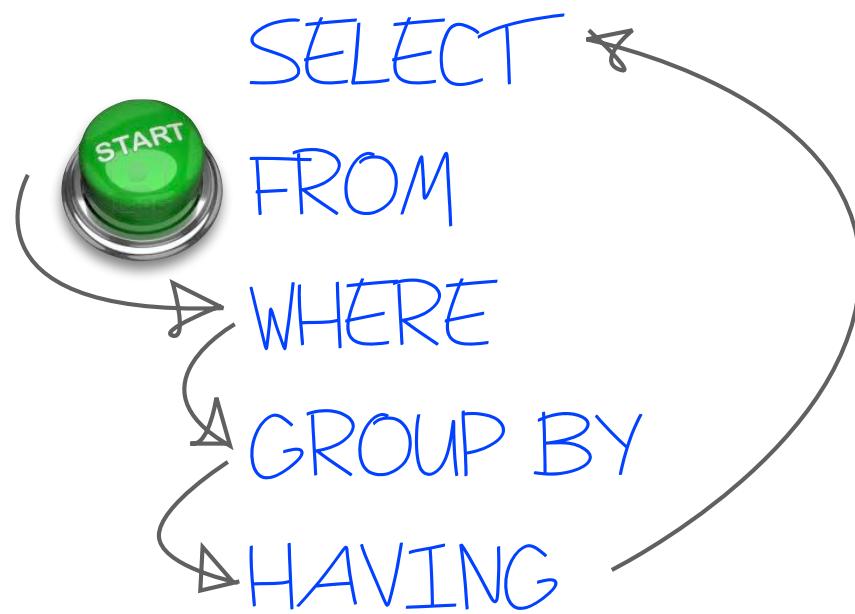
```
SELECT Specializzazione  
FROM Medico  
WHERE Citta = 'Pisa'  
GROUP BY Specializzazione  
HAVING COUNT(*) > 2;
```



Si contano i record all'interno di ciascun gruppo. Ogni record è relativo a un medico.

# Ordine di esecuzione

---



# Common Table Expression (CTE)

---

si scrivono, separati da virgola, subito prima della query che li usa,  
tramite la keyword `WITH`

Sono **result set dotati di identificatore** che possono essere usati  
prima di una query per costruire risultati intermedi.


sono parti di codice i cui risultati sono stoccati in memoria, nominati,  
e usati dalla query immediatamente dopo

# Sintassi di WITH per dichiarare CTEs

---

**WITH**

*name1 AS (query1)*  
*, name2 AS (query2)*  
*,*  
*...*  
*, nameN AS (queryN)*

*query finale;*

una query finale può usare più CTEs  
che vanno dichiarate prima di essa,  
separate da virgola



usa nel *from* le CTEs *name1, name2, ..., nameN*  
dichiarate immediatamente prima

# Esempio di uso di CTE (rif. slide 40)

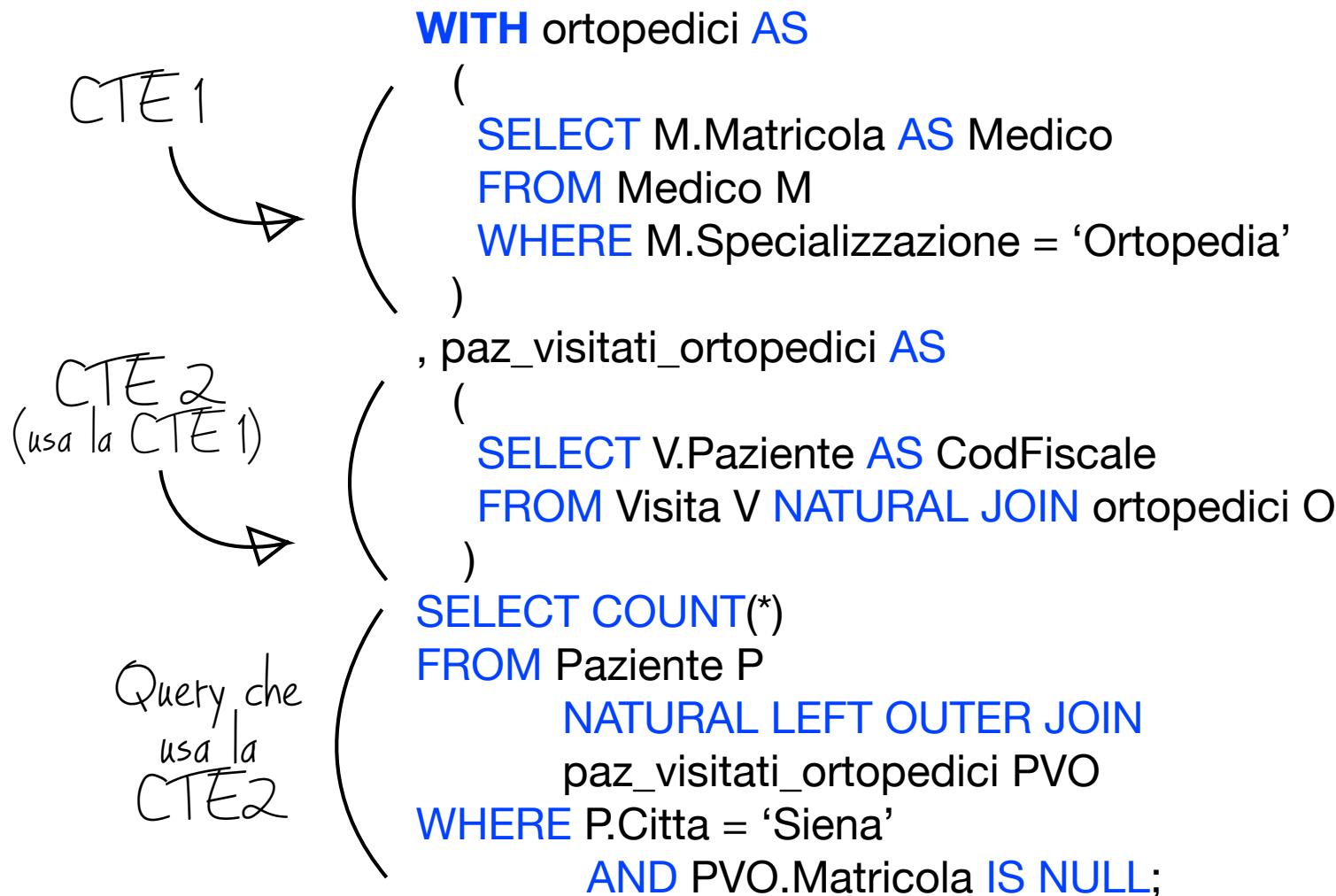
Indicare il numero di pazienti di Siena, mai visitati da ortopedici

WITH pazienti\_visitati\_ortopedici AS  
(  
    SELECT V.Paziente AS CodFiscale  
        FROM Visita V  
            INNER JOIN  
                Medico M ON V.Medico = M.Matricola  
            WHERE M.Specializzazione = 'Ortopedia'  
                AND P.Citta = 'Siena'

Query  
che usa  
la CTE  
)  
SELECT COUNT(\*)  
    FROM Paziente P  
        NATURAL RIGHT OUTER JOIN  
            pazienti\_visitati\_ortopedici PVO  
            WHERE PVO.CodFiscale IS NULL;

# Stesso esempio, più CTE

Indicare il numero di pazienti di Siena, mai visitati da ortopedici



# Esercizi

---

**Prima di provare a risolvere gli esercizi seguenti consiglio di rivedere tutti gli esempi mostrati in questo pacchetto di slide e provare a risolvere prima quelli.**

1. Considerata ogni specializzazione, indicarne il nome e l'incasso degli ultimi due anni.
2. Indicare le specializzazioni aventi medici della stessa città.
3. Indicare codice fiscale, nome, cognome ed età del paziente più anziano della clinica, e il numero di volte da cui è stato visitato da ogni medico.
4. Indicare la matricola dei medici che hanno effettuato più del 20% delle visite annue della loro specializzazione in almeno due anni fra il 2010 e il 2020. [Suggerimento: nel select, è possibile inserire espressioni (quindi si possono usare +,-,\*,/). Per esempio, se voglio restituire il reddito annuale di tutti i pazienti della clinica, posso scrivere:  
**SELECT P.Reddito\*12 FROM Paziente P;**]
5. Fra tutte le città da cui provengono più di tre pazienti con reddito superiore a 1000 Euro, indicare quelle da cui provengono almeno due pazienti che sono stati visitati più di una volta al mese, nel corso degli ultimi 10 anni.
6. Rivedere gli esercizi della scorsa settimana e trovare quelli che si possono risolvere usando il raggruppamento. Una volta individuati, risolverli con il raggruppamento.