

ΣΥΓΚΕΝΤΡΩΤΙΚΕΣ ΣΥΝΑΡΤΗΣΕΙΣ

GROUP BY - HAVING

Παραδείγματα

- Συνολικό πλήθος έργων που προμηθεύει ο προμηθευτής S1
`select COUNT(J#) from SPJ where S# = 'S1'`
- Μετά την επιλογή, δεν χρειάζεται ομαδοποίηση, όλος ο πίνακας που μένει είναι μια ομάδα
- Είναι όμως η σωστή απάντηση ?

Παραδείγματα

S#	P#	J#	POSOT
S1	P1	J2	300
S1	P2	J1	400
S2	P3	J2	200
S2	P2	J1	150
S3	P4	J3	200
S1	P2	J1	230

- Τα έργα του S1 είναι το J1 και το J2
- Αν μετρήσουμε τις γραμμές, βρίσκουμε 3.
- Λύση: DISTINCT – μετράμε κάθε τιμή μόνο μία φορά
- Τελικά: `select COUNT(DISTINCT J#) from SPJ where S# = 'S1'`

Παραδείγματα

- Συνολική ποσότητα για κάθε υλικό που προμηθεύει ο S1

```
select SUM(posot) from SPJ where S# = 'S1'  
GROUP BY P#
```
- SELECT: επιδρά σε γραμμές πίνακα
- Μπορούμε να κάνουμε και υπολογισμούς
- Παράδειγμα: Άθροισμα των ποσών των καταθέσεων σε κάθε κατάστημα

Παραδείγματα

- Για κάθε υλικό που παρέχεται σε κάποιο έργο, κωδικός υλικού, κωδικός έργου Συνολική ποσότητα για κάθε υλικό που προμηθεύει ο S1

```
select SUM(posot) from SPJ where S# = 'S1'  
GROUP BY P#
```

- SELECT: επιδρά σε γραμμές πίνακα
- Μπορούμε να κάνουμε και υπολογισμούς
- Παράδειγμα: Άθροισμα των ποσών των καταθέσεων σε κάθε κατάστημα

Ομαδοποίηση

- «για κάθε» κατάσταση → δημιουργία ομάδων

- ✓ Ομάδα «Κέντρο»: $100+250 = 350$

- ✓ Ομάδα «Ομόνοια»: **180**

- ✓ Ομάδα «Πύργος»: **310**

- Άρα αποτέλεσμα:

Κέντρο	350
Ομόνοια	180
Πύργος	310

Συγκεντρωτικές (aggregate) συναρτήσεις

- Η SQL υποστηρίζει:
 - ✓ Άθροισμα SUM()
 - ✓ Μέσος όρος AVG()
 - ✓ Ελάχιστο MIN()
 - ✓ Μέγιστο MAX()
 - ✓ Μετρητής COUNT()
- Πώς γίνονται οι ομάδες?

Ο όρος GROUP BY

- Σχηματίζω **ομάδες** με βάση τις **τιμές κάποιου πεδίου**
- Θέλω άθροισμα καταθέσεων **για κάθε κατάσταση**

SELECT

FROM Deposit

GROUP BY Br-name

- Σχηματίζω ομάδες και μετά εκτελώ το **SELECT**

SELECT sum(balance)		(Κέντρο)	350
FROM Deposit	Αποτέλεσμα	(Ομόνοια)	180
GROUP BY Br-name		(Πύργος)	310

Για κάθε κατάσταση, δείξε ένα άθροισμα

SELECT συμβατό με GROUP BY (1)

- Προσοχή: όλη η εντολή πρέπει να συμφωνεί με την ομαδοποίηση
→ **δεν μπορώ να έχω τα πάντα στο SELECT**
- **Ερώτημα:** Τι μπορώ να έχω στο SELECT?
- **Σίγουρα συναρτήσεις – αφορούν σε ομάδες**
- Τι άλλο? Τι από όλα τα πεδία είναι συμβατό με το group by?
(σκεφτείτε)

SELECT συμβατό με GROUP BY (2)

Μπορώ να έχω το παρακάτω?

```
SELECT cust-name, SUM(balance)
      FROM Deposit
      GROUP BY Br-name
```

Το παρακάτω?

```
SELECT br-name, SUM(balance)
      FROM Deposit
      GROUP BY Br-name
```

SELECT συμβατό με GROUP BY (3)

- Το cust-name μπορεί να ομαδοποιηθεί με βάση το Br-name?

ΌΧΙ!

- Το br-name μπορεί να ομαδοποιηθεί με βάση το Br-name?

ΝΑΙ!

Συμπέρασμα: μόνο το πεδίο ομαδοποίησης και συναρτήσεις

Συναρτήσεις χωρίς GROUP BY (1)

- Μπορώ να μην βάλω group by αν έχω συναρτήσεις?

```
SELECT AVG (amount)  
FROM Borrow
```

- Όλος ο πίνακας μια ομάδα
- Υπολογίζεται ένα μέγεθος για όλο τον πίνακα
- Χρήση: μπορώ να το χρησιμοποιώ όπου χρειαζομαι αυτό το μέγεθος
- Συνέπεια: χωρίς Group By, μόνο συναρτήσεις στο SELECT

Συναρτήσεις χωρίς GROUP BY (1)

- Μπορώ να μην βάλω group by αν έχω συναρτήσεις?

```
SELECT AVG (amount)  
FROM Borrow
```

- Όλος ο πίνακας μια ομάδα
- Υπολογίζεται ένα μέγεθος για όλο τον πίνακα
- Χρήση: μπορώ να το χρησιμοποιώ όπου χρειαζομαι αυτό το μέγεθος
- Συνέπεια: χωρίς Group By, μόνο συναρτήσεις στο SELECT

Συνθήκες

- Το WHERE εφαρμόζεται αν χρειάζεται
- Δεν επηρεάζεται από το GROUP BY
- Παράδειγμα: πόσους λογαριασμούς έχουν ο Πέτρου και ο Παύλου
SELECT COUNT(Cust-name) , Cust-name
FROM Deposit
WHERE Cust-name IN ('Πέτρου', 'Παύλου')
ORDER BY Cust-name
GROUP BY Cust-name
- Πρώτα σχηματίζεται ο πίνακας με την επιλογή γραμμών
- Μετά **στον νέο πίνακα** γίνεται η ομαδοποίηση
- Τα ίδια ισχύουν για το ORDER BY

Τι λαμβάνεται υπόψη?

- Τι βγάζει η εντολή `SELECT Br-name, AVG(balance) FROM Deposit GROUP BY Br-name`

Acc-no	Br-name	Balance	Cust-name
123	Κέντρο	100	...
522	Κέντρο	250	...
4512	<u>Ομονοια</u>	180	...
4521	Πύργος	310	...
6738	Πύργος	1000	...
5969	Κέντρο		...
2929	Ομόνοια		...

Τι λαμβάνεται υπόψη?

- Η τιμή 'Κέντρο' έχει 3 γραμμές, **αλλά 2 τιμές** (μία κενή)
- Άρα μέσος όρος $(100+250) / 2 = 175$

Acc-no	Br-name	Balance	Cust-name
123	Κέντρο	100	...
522	Κέντρο	250	...
4512	<u>Ομονοια</u>	180	...
4521	Πύργος	310	...
6738	Πύργος	1000	...
5969	Κέντρο		...
2929	Ομόνοια		...

- Δεν λαμβάνονται υπ' όψη οι null τιμές

Μετρητής (COUNT)

- Η count() μετράει γραμμές
- Πόσες παραγγελίες έχει κάθε έργο?

```
SELECT J#, COUNT(J#)
```

```
FROM SPJ
```

```
GROUP BY J#
```

- Έχει διαφορά αν βάλουμε COUNT(S#) ?
- Απάντηση: ΟΧΙ! Μετράμε τιμές του πεδίου, δηλαδή γραμμές
- Γι' αυτό υπάρχει και η δυνατότητα COUNT (*)

Μετρητής (COUNT)

- Προσοχή: τα κενά δεν συμμετέχουν σε καμία συνάρτηση
- Γι' αυτό έχει σημασία ποιο πεδίο επιλέγουμε στο COUNT
- Αν μας ενδιαφέρουν μόνο οι γραμμές, μετράμε το πρωτεύον κλειδί ή βάζουμε *
- Κάποιες φορές η διατύπωση του ερωτήματος μας δείχνει ποιο πεδίο να επιλέξουμε
- Προσοχή: το COUNT είναι εντελώς διαφορετικό από το SUM

HAVING

- Πώς να επιλέξω ομάδες?
- Παράδειγμα: Πόσες παραγγελίες έχει κάθε έργο, αλλά μόνο για έργα με πάνω από 5 παραγγελίες
- Το WHERE επιλέγει γραμμές → δεν είναι κατάλληλο
- Τη δουλειά αυτή κάνει το HAVING

HAVING

```
SELECT J#, COUNT (J#)  
FROM SPJ  
GROUP BY J#  
HAVING COUNT(J#) >5
```

- Μόνο συναρτήσεις έχουν νόημα στη συνθήκη του HAVING (θυμηθείτε: συμβατή με το GROUP BY)
- Μπορεί και πάνω από μία συνθήκη: HAVING COUNT(J#) >5 AND COUNT(J#) <8