



ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΙΓΑΙΟΥ

Βάσεις Δεδομένων II

Εισαγωγή

Μανώλης Μαραγκουδάκης

Τμήμα Μηχανικών Πληροφοριακών και Επικοινωνιακών Συστημάτων



Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο



ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ & ΘΡΗΣΚΕΥΜΑΤΩΝ, ΠΟΛΙΤΙΣΜΟΥ & ΑΘΛΗΤΙΣΜΟΥ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης



Άδειες Χρήσης

- Το παρόν εκπαιδευτικό υλικό υπόκειται σε άδειες χρήσης Creative Commons.
- Για εκπαιδευτικό υλικό, όπως εικόνες, που υπόκειται σε άλλου τύπου άδειας χρήσης, η άδεια χρήσης αναφέρεται ρητώς.



Χρηματοδότηση

- Το παρόν εκπαιδευτικό υλικό έχει αναπτυχθεί στα πλαίσια του εκπαιδευτικού έργου του διδάσκοντα.
- Το έργο «**Ανοικτά Ακαδημαϊκά Μαθήματα στο Πανεπιστήμιο Αιγαίου**» έχει χρηματοδοτήσει μόνο τη αναδιαμόρφωση του εκπαιδευτικού υλικού.
- Το έργο υλοποιείται στο πλαίσιο του Επιχειρησιακού Προγράμματος «Εκπαίδευση και Δια Βίου Μάθηση» και συγχρηματοδοτείται από την Ευρωπαϊκή Ένωση (Ευρωπαϊκό Κοινωνικό Ταμείο) και από εθνικούς πόρους.



Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο



ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ & ΘΡΗΣΚΕΥΜΑΤΩΝ, ΠΟΛΙΤΙΣΜΟΥ & ΑΘΛΗΤΙΣΜΟΥ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης



ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ

Πανεπιστήμιο Αιγαίου-
Τμήμα Μηχανικών
Πληροφοριακών και
Επικοινωνιακών Συστημάτων



Βάσεις Δεδομένων II

ΕΙΣΑΓΩΓΗ: Παρουσίαση στόχων μαθήματος
Μανώλης Μαραγκουδάκης

www.icsd.aegean.gr/mmarag

Τι θα μας απασχολήσει στο μάθημα

- **ΜΕΡΟΣ I: Εσωτερική δομή ενός ΣΔΒΔ**
 - Επεξεργασία και βελτιστοποίηση ερωτήσεων
 - Διαχείριση συναλλαγών και έλεγχος ταυτοχρονισμού
 - Ανάνηψη από αποτυχίες
 - Φυσική σχεδίαση και ρύθμιση μιας βάσης δεδομένων
 - Διαχείριση χρηστών και ασφάλεια
- **ΜΕΡΟΣ II: Εφαρμογές γύρω από βάσεις δεδομένων**
 - Αποθήκες δεδομένων
 - Εξόρυξη γνώσης

Τι θα μας απασχολήσει στο μάθημα

- Μέρος III: Ειδικά θέματα βάσεων δεδομένων
 - Αντικειμενοστρεφείς και αντικειμενο-σχεσιακές βάσεις δεδομένων
 - Το πρότυπο SQL-1999
 - Η γλώσσα ερωτήσεων Datalog
 - Το επεκτεταμένο μοντέλο Οντοτήτων-Συσχετίσεων
 - Χωροχρονικές, διαδικτυακές και ενεργές βάσεις δεδομένων.

[αν προλάβουμε ...]

Για να περάσω το μάθημα...

- Βαθμολογία :
 - **Διαγώνισμα:** 50% [βαθμός >4]
 - **Εργασία:** 30% [βαθμός >=5, υποχρεωτική]
 - Πιθανό bonus ως 10% σε καλές εργασίες...
 - **Εργαστηριακές Ασκήσεις-Εργασίες:** 20% [υποχρεωτικές]
 - **Θεωρητικές Ασκήσεις:** 10% [bonus]

Προγραμματιστική Εργασία

- Η προγραμματιστική άσκηση για το μάθημα είναι **υποχρεωτική**
- Αφορά τη σχεδίαση, υλοποίηση και ρύθμιση ενός **ολοκληρωμένου πληροφοριακού συστήματος**
 - προδιαγραφές και ανάλυση
 - κατασκευή βάσης δεδομένων
 - κατασκευή διαπροσωπείας
 - ρύθμιση λειτουργίας
- **Είναι βασικό η εργασία να ολοκληρωθεί πλήρως!!**

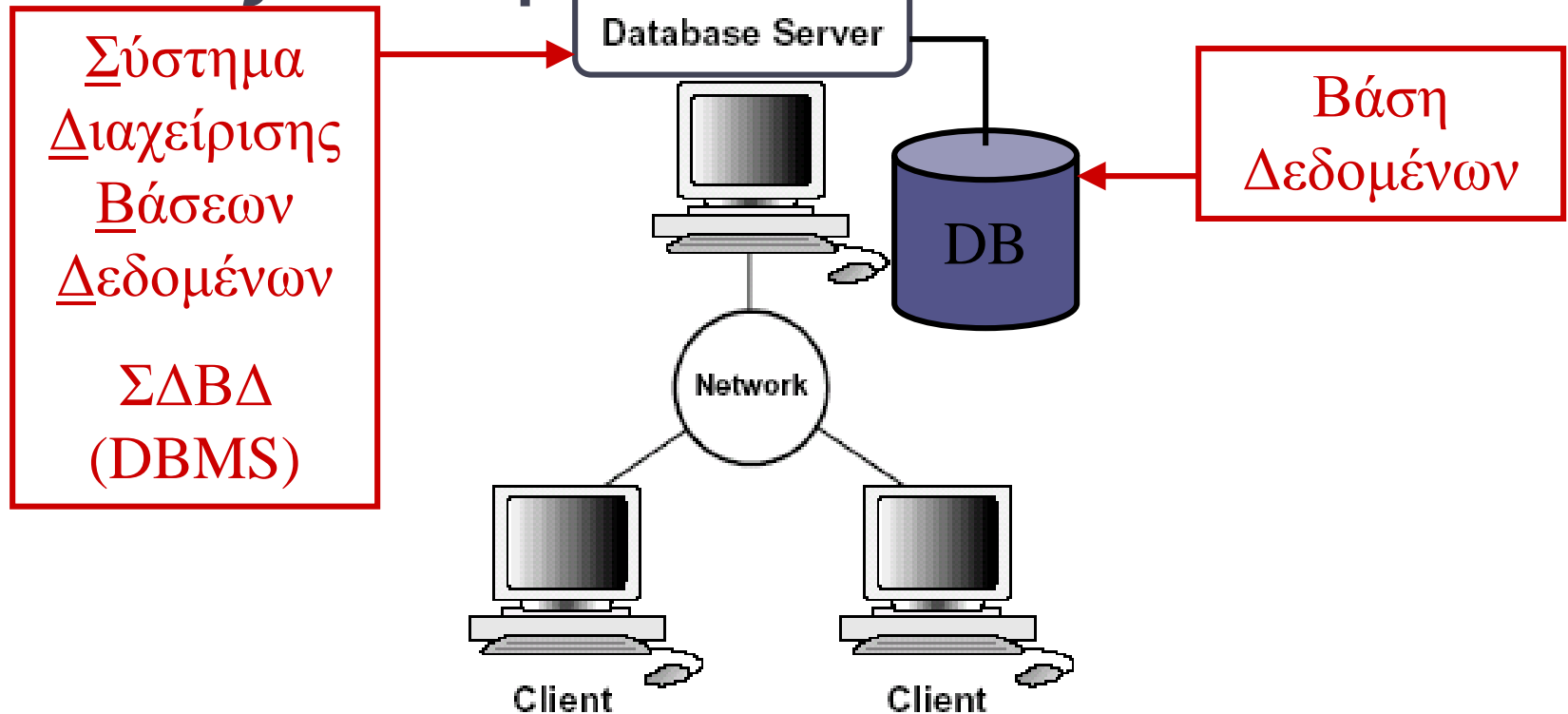
Ασκήσεις εξάσκησης (*bonus*)

- Θα δοθούν **2 ή 3 σειρές ασκήσεων** (ανάλογα με το χρόνο) για να σας βοηθήσουν με τις έννοιες του μαθήματος.
 - *Για να βοηθήσουν, θα πρέπει να τις κάνετε και όχι να τις αντιγράψετε...*

Βιβλιογραφία

- Διδακτικά Βιβλία:
 - «Θεμελιώδεις Αρχές Συστημάτων Βάσεων Δεδομένων», Τόμος Α' & Β', R. Elmasri & S.B. Navathe (μετάφραση Μ. Χατζόπουλος), Εκδόσεις Δίαυλος, 1998
 - «Συστήματα Διαχείρισης Βάσεων Δεδομένων», Τόμος Β', R. Ramakrishnan & J. Gehrke, (μετάφραση Δ. Δέρβος, Α. Ευαγγελίδης) Εκδόσεις Τζιόλα 2002.
 - «Συστήματα Βάσεων Δεδομένων - Η πλήρης θεωρία των βάσεων δεδομένων», A. Silberschatz, H.F. Korth, και S. Sudarshan (μετάφραση Μ. Γκλαβά), 4η έκδοση, Εκδόσεις Γκιούρδα, 2004

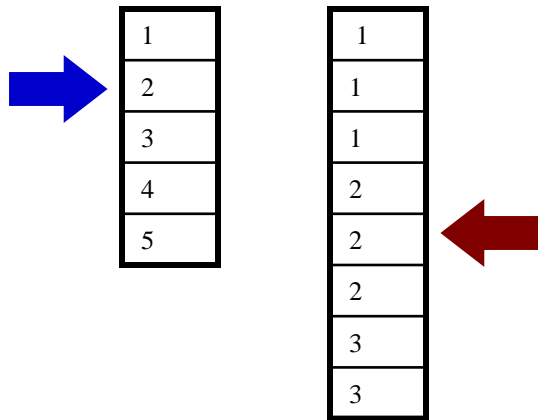
Βάσεις Δεδομένων και ΣΔΒΔ



Βάσεις Δεδομένων και ΣΔΒΔ

- Μια **Βάση Δεδομένων** είναι μια οργανωμένη συλλογή θεματικά συναφών δεδομένων (συνήθως σε μεγάλη κλίμακα)
- Ένα **Σύστημα Διαχείρισης Βάσεων Δεδομένων (DataBase Management System – DBMS)** είναι ένα πακέτο λογισμικού που σκοπό έχει τη διαχείριση βάσεων δεδομένων

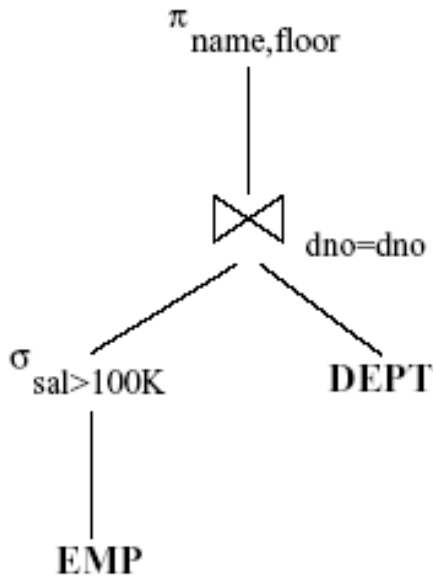
Επεξεργασία Ερωτήσεων



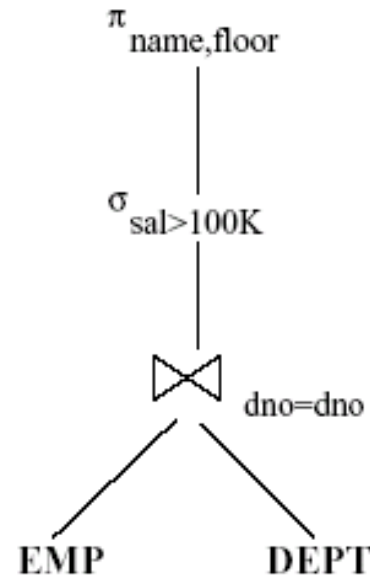
```
SELECT E.ename, D.mgr
FROM Emp E, Dept D
WHERE D.dname='Toys' AND
E.dno=D.dno
```

Ταξινομήσε τα Emp, Dept με βάση το πεδίο Dno
 Για κάθε $D \in \text{Dept}$ [$\& Dname='Toys'$]
 Για κάθε $E \in \text{Emp}$ με ίδιο Dno
 Επέστρεψε E.ename, D.mgr

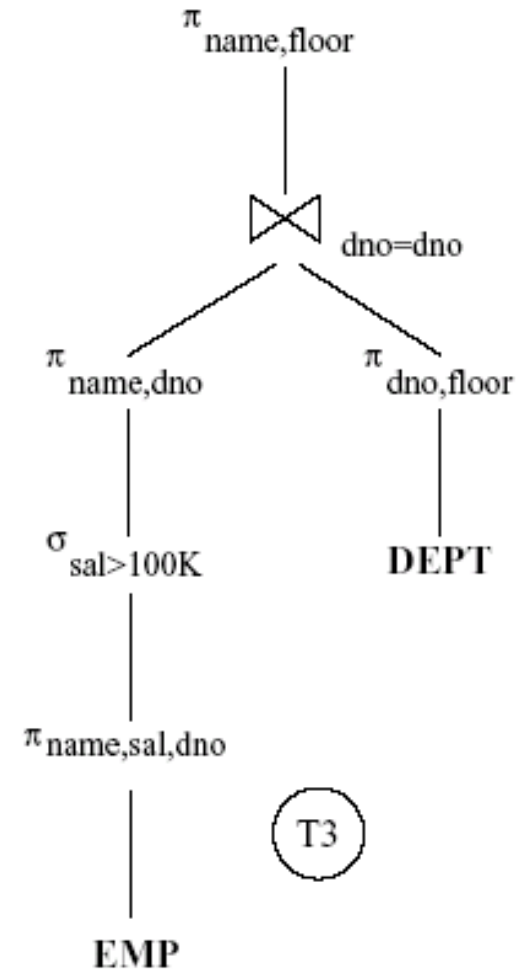
Παράδειγμα ισοδύναμων, εναλλακτικών πλάνων



T1



T2



T3

```

select name, floor
from emp, dept
where emp.dno =
dept.dno
and sal > 100K

```

Ρύθμιση της σχεδίασης

Emp

<u>Eno</u>	Ename	Dno	Sal	Hobby
103	Demis	1	10K	Music
202	Vassilis	2	8K	Food
208	Ilia	2	10K	Coffee
123	Theo	2	5K	sports

Το σχήμα των πινάκων είναι ίδιο

Η ένωση των πλειάδων δίνει τις πλειάδες του αρχικού πίνακα

Emp_Rich

Sal >= 10K

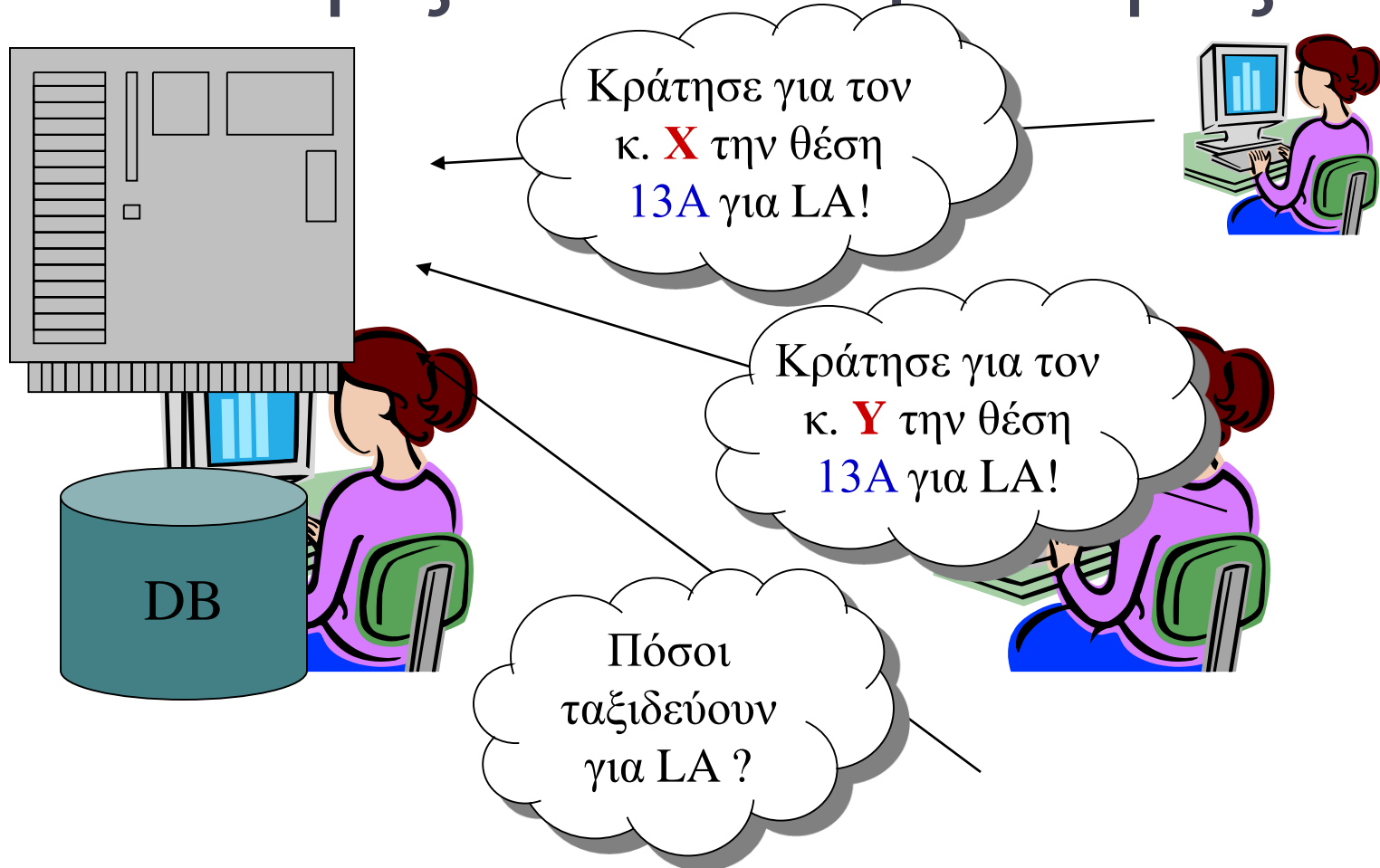
<u>Eno</u>	Ename	Dno	Sal	Hobby
103	Demis	1	10K	Music
208	Ilia	2	10K	Coffee

Emp_Poor

Sal < 10K

<u>Eno</u>	Ename	Dno	Sal	Hobby
202	Vassilis	2	8K	Food
123	Theo	2	5K	sports

Συναλλαγές & Ταυτοχρονισμός



Συναλλαγή

- **Συναλλαγή** είναι
 - μια σειρά από ενέργειες, οι οποίες
 - διαβάζουν ή γράφουν
 - αντικείμενα της βάσης
- στα αγγλικά “transaction”
- σειρά: διατεταγμένο σύνολο, **λίστα**

Σημείωση: στην θεωρία [της πληροφορικής γενικά, και των ΒΔ ειδικά], οι λέξεις του τύπου «σειρά» έχουν σημασία ... (π.χ., σειρά \neq σύνολο)

Παράδειγμα συναλλαγής

T_0 : μεταφορά 50€ από
το λογαριασμό A στο
λογαριασμό B

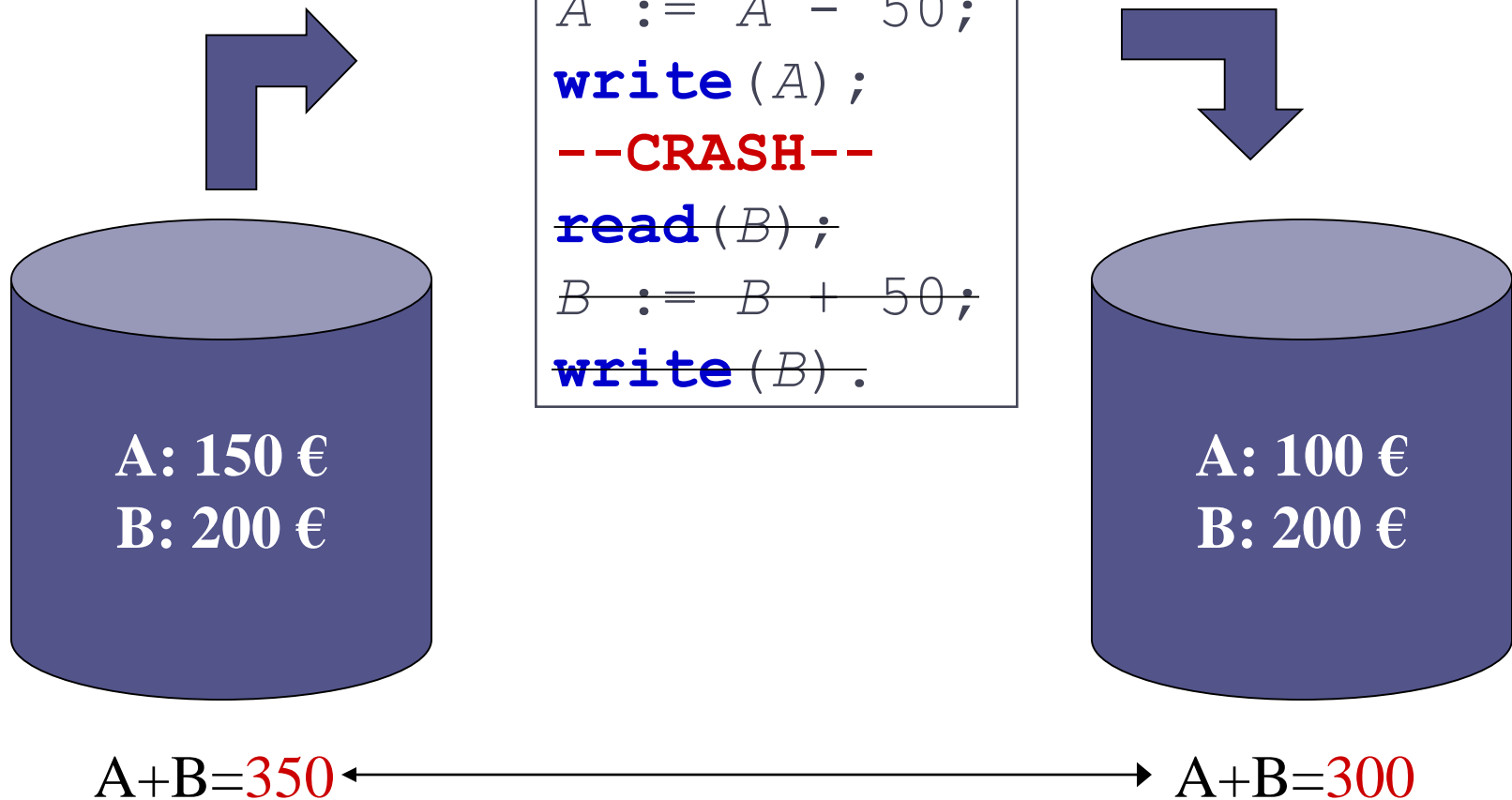
```
read (A) ;  
A := A - 50 ;  
write (A) ;  
read (B) ;  
B := B + 50 ;  
write (B) .
```

Προβληματισμός...

- Δύο είναι τα βασικά προβλήματα με τις συναλλαγές:
 - Τι θα γίνει αν κατά τη διάρκεια της εκτέλεσης, πέσει το σύστημα?
 - Τι θα γίνει αν δύο συναλλαγές επιχειρούν να μεταβάλλουν το ίδιο αντικείμενο ταυτοχρόνως?

*Ορολογία: Στο εξής το σύστημα δεν «πέφτει»,
αλλά «αποτυγχάνει»*

Αποτυχία

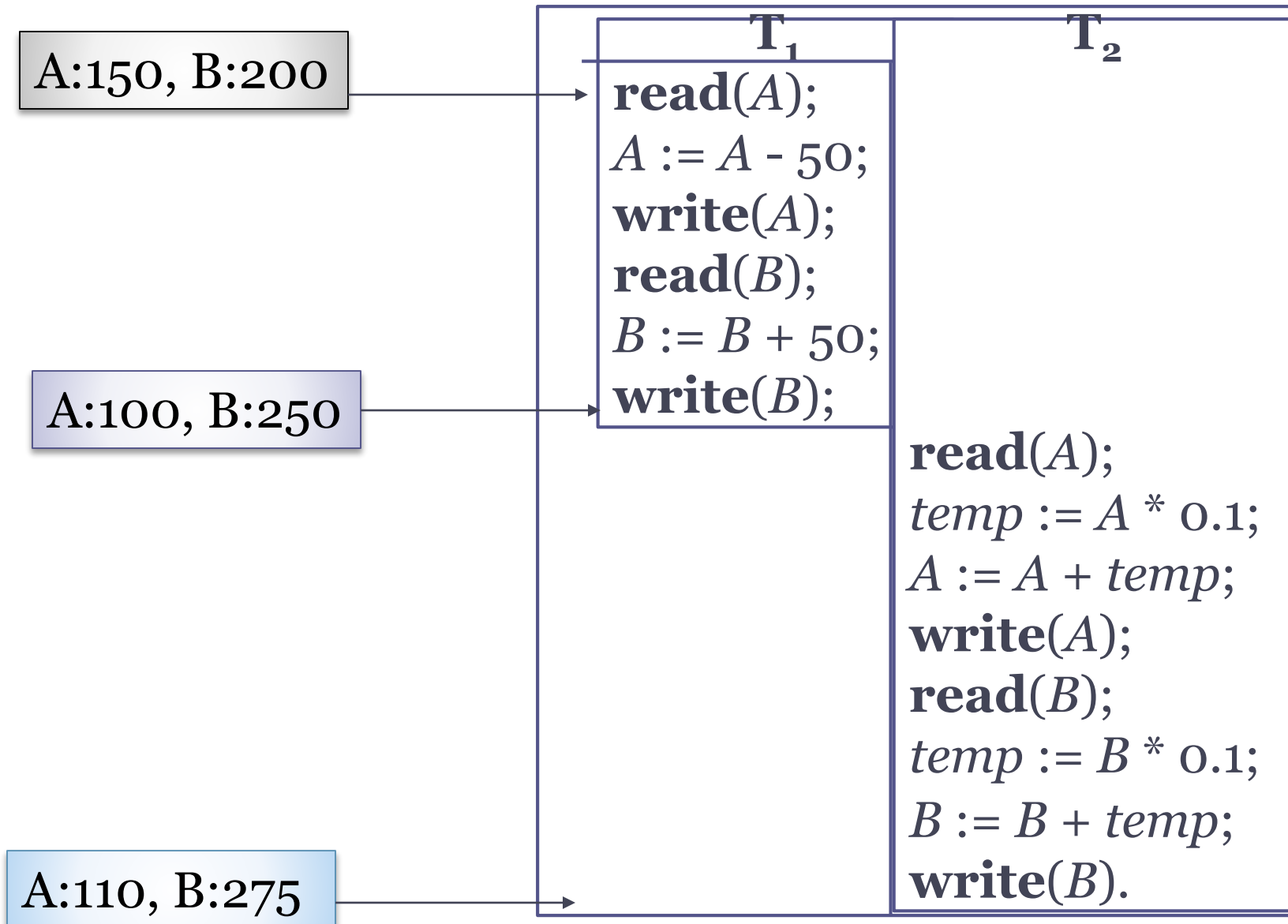


Ταυτοχρονισμός

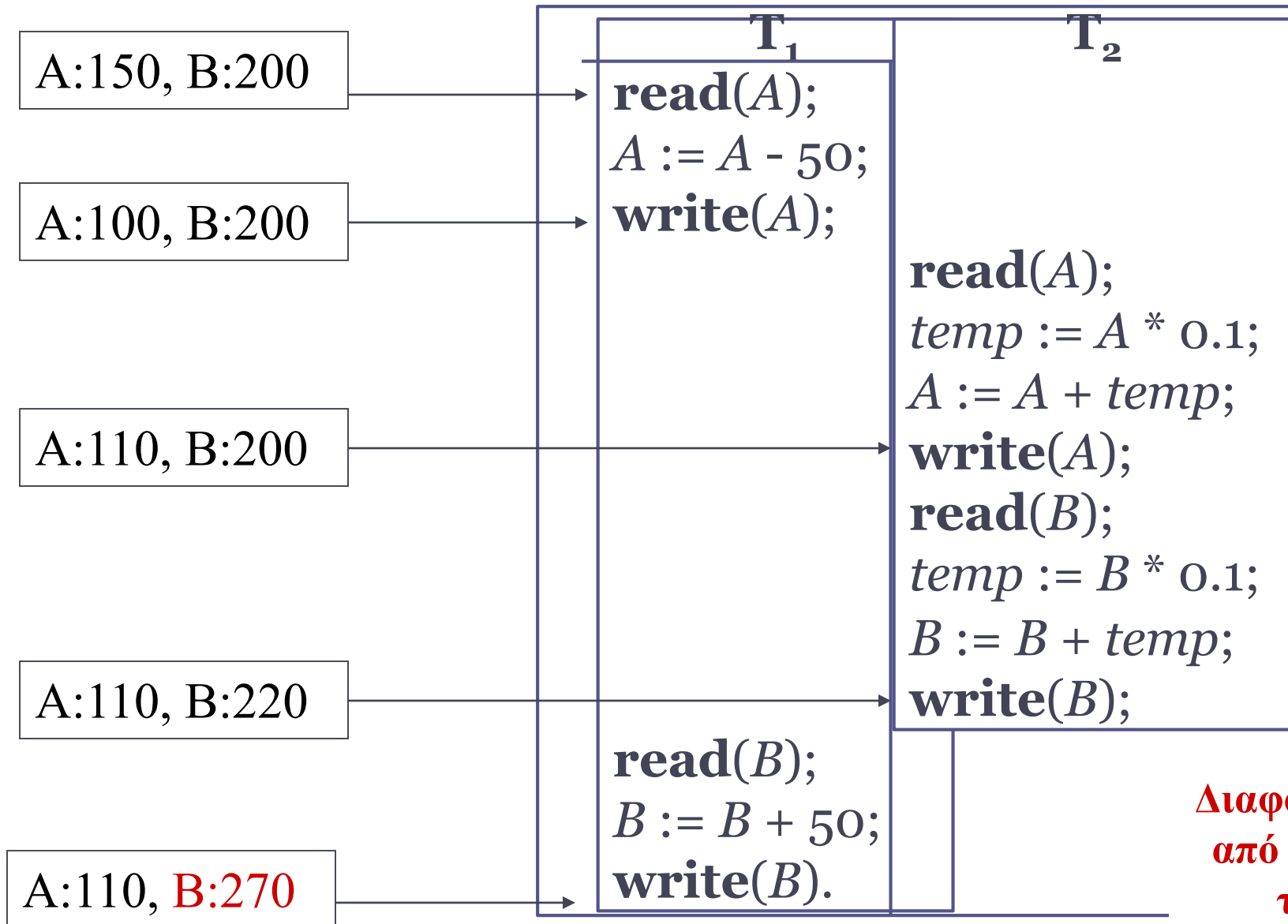
Έστω δύο συναλλαγές που τρέχουν
σειριακά:

- T1: μεταφέρει 50 € από τον A στον B
- T2: κάνει αύξηση στον A και στο B
κατά 10%

Ταυτοχρονισμός σειριακά



Ταυτοχρονισμός με πολυπλεξία συναλλαγών



**Διαφορετικό
από το 275
του
σειριακού!!!**

Αστοχίες του συστήματος

- Κακό πρόγραμμα (με λάθη, δηλ.)

⇒ Αστοχία **δοσοληψίας** (αδιέξοδο, abort από τον χρήστη, κλπ)

⇒ Αστοχία του **συστήματος** (πτώση ρεύματος, αδιέξοδο λειτουργικού συστήματος)

- Αστοχία **υλικού** (καταστροφή σκληρού δίσκου)

Failure: αστοχία ή αποτυχία

Σε περίπτωση αποτυχίας ...

- Σκοπός είναι να διατηρήσουμε τη συνέπεια του συστήματος.
- Πρέπει να επαναλάβουμε (**REDO**) όλες τις δοσοληψίες που έκαναν commit
- Πρέπει να αναιρέσουμε (**UNDO**) όσες παρενέργειες επέφεραν οι δοσοληψίες που δεν πρόλαβαν να κάνουν commit

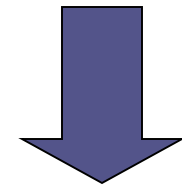
Log (Ιστορικό)

- **Log** (Ιστορικό/Ίχνος/Ημερολόγιο): ένα αρχείο στο σκληρό δίσκο που καταγράφει όλη την ιστορία των ενεργειών που εκτελέστηκαν από το DBMS
- **Ενέργειες:**
 - BOT/EOT (Begin/End Of Transaction)
 - INS/UPD/DEL (δηλ. write) σε μια εγγραφή
 - COMMIT/ABORT μια δοσοληψία
 - UNDO/REDO μια ενέργεια εγγραφής (write)

Log Record τύπου “Write”

- **LSN** (μονοσήμαντο αναγνωριστικό του log entry)
- **Transaction ID (TID)**
- Σελίδα που κάνουμε update
- **Offset στην εν λόγω σελίδα**
- Μήκος σε bytes που αλλάζουμε
- **Παλιά τιμή**
- **Νέα τιμή**

```
T1: UPDATE EMP  
SET ID = 30  
WHERE ID = 3
```



```
Log Entry: LSN12, T1, PAGE32, 0xFFF32, 8, 3, 30
```

Write Ahead Log

***Προτού γράφεις οτιδήποτε στη ΒΔ,
καταχώρησε την αντίστοιχη εγγραφή στο
log***

- Για να γράφεις μια updated σελίδα από το buffer πίσω στο δίσκο, πρέπει στο log (στο δίσκο) να έχουν περαστεί οι παλιές τιμές για τα records της
- Για να κάνεις commit μια δοσοληψία πρέπει στο log (στο δίσκο) να έχουν γραφτεί όλες οι σχετικές log records

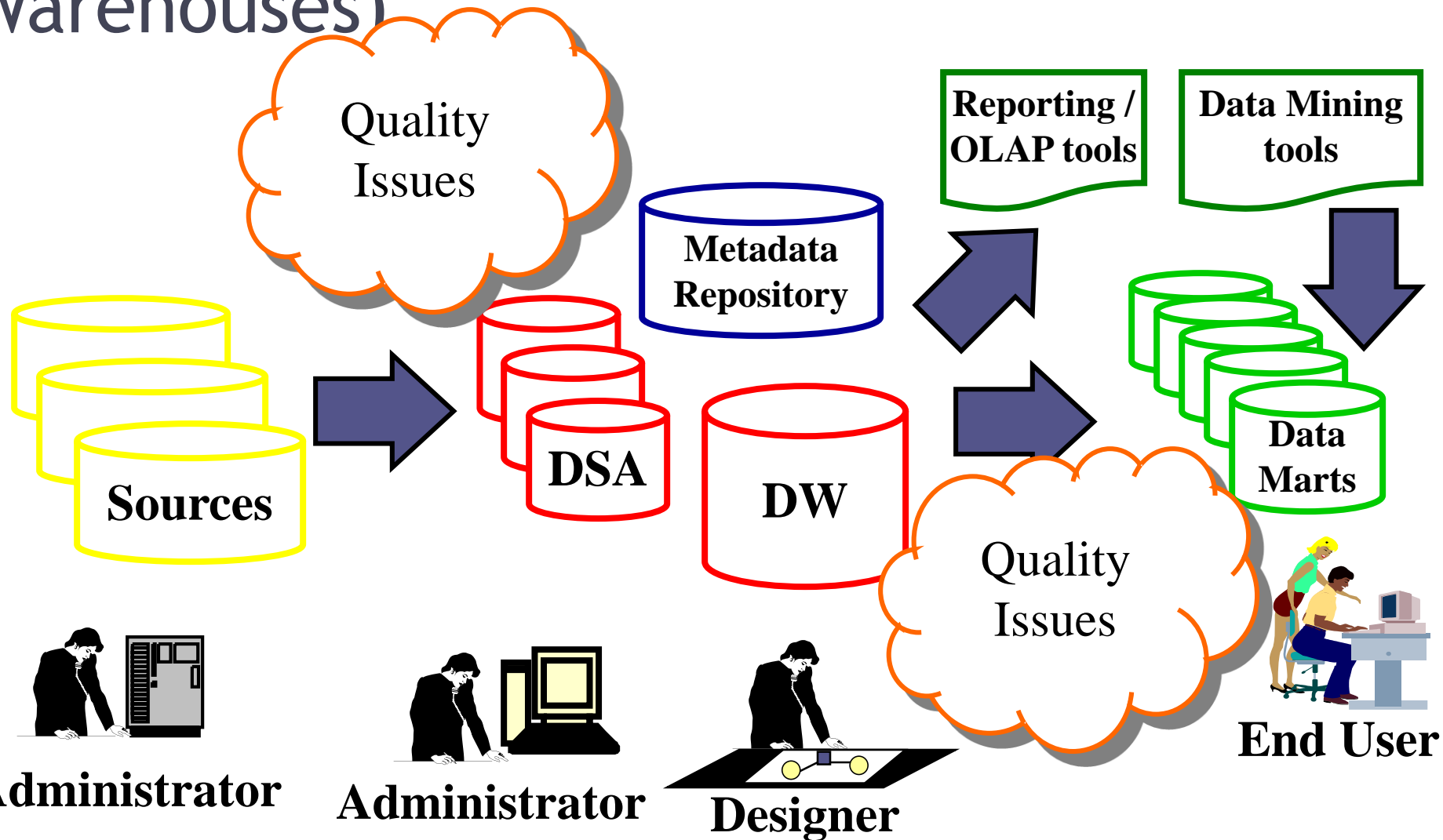
Checkpoints - Σημεία ελέγχου

- Περιοδικά, το σύστημα κάνει τις εξής ενέργειες:
 - Σταματά κάθε άλλη ενέργεια
 - Καταγράφει το σύνολο των ενεργών δοσοληψιών
 - Γράφει (**flush**) όλους τους buffers με log records, στο δίσκο
 - Γράφει (**flush**) όλους τους buffers με records της ΒΔ, στο δίσκο
 - Γράφει στο log μια εγγραφή **CHK** (checkpoint)

Ανάνηψη αν έχουμε WAL

- Έστω ότι το σύστημα αποτυγχάνει και πρέπει να το επαναφέρουμε (ήτοι, να επαναφέρουμε τη ΒΔ σε συνεπή μορφή).
- Η διαδικασία αυτή ονομάζεται **ανάνηψη (recovery)** ή **ανάκαμψη** ή **επαναφορά**
- Αν έχουμε χρησιμοποιήσει **WAL** κατά την κανονική λειτουργία του συστήματος, η ανάνηψη έχει **3 φάσεις**:
 1. Ανάλυση
 2. UNDO
 3. REDO

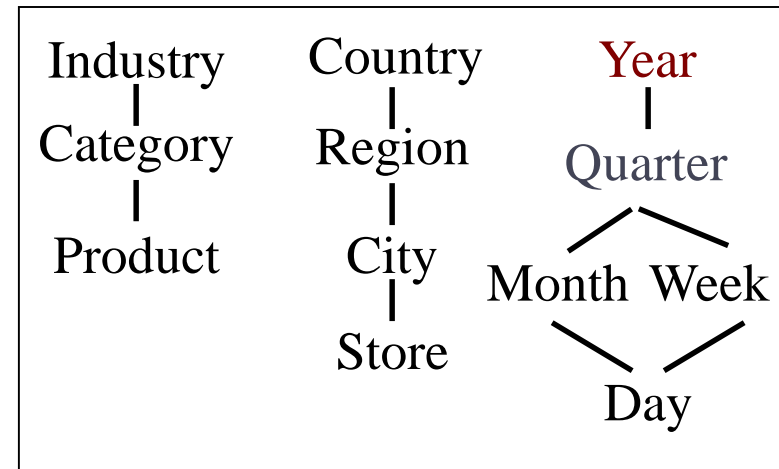
Αποθήκες Δεδομένων (Data Warehouses)



Roll up

Sales volume			
	Products	Store1	Store2
Q1	Electronics	\$5,2	\$5,6
	Toys	\$1,9	\$1,4
	Clothing	\$2,3	\$2,6
	Cosmetics	\$1,1	\$1,1
Q2	Electronics	\$8,9	\$7,2
	Toys	\$0,75	\$0,4
	Clothing	\$4,6	\$4,6
	Cosmetics	\$1,5	\$0,5

Χρόνος: Επίπεδο Quarter

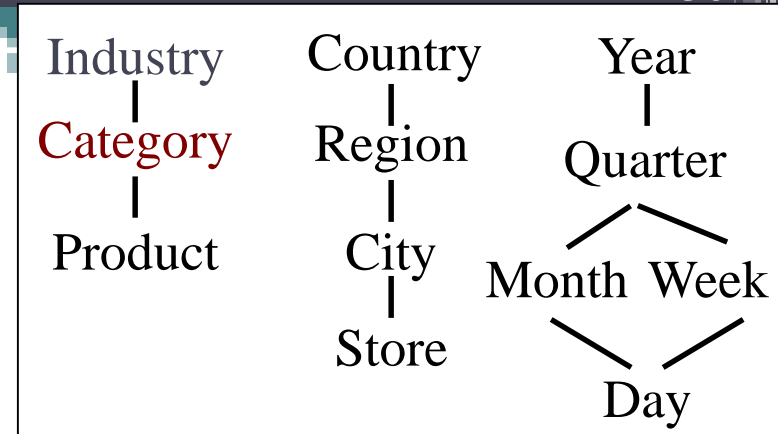


Sales volume			
	Products	Store1	Store2
Year 1996	Electronics	\$14,1	\$12,8
	Toys	\$2,65	\$1,8
	Clothing	\$6,9	\$7,2
	Cosmetics	\$2,6	\$1,6

Χρόνος: Επίπεδο Year

SUM(Sales volumes)

Drill down



Sales volume

		Products	Store1	Store2
Q1		Electronics	\$5,2	\$5,6
		Toys	\$1,9	\$1,4
		Clothing	\$2,3	\$2,6
		Cosmetics	\$1,1	\$1,1
Q2		Electronics	\$8,9	\$7,2
		Toys	\$0,75	\$0,4
		Clothing	\$4,6	\$4,6
		Cosmetics	\$1,5	\$0,5

Item: Επίπεδο Industry

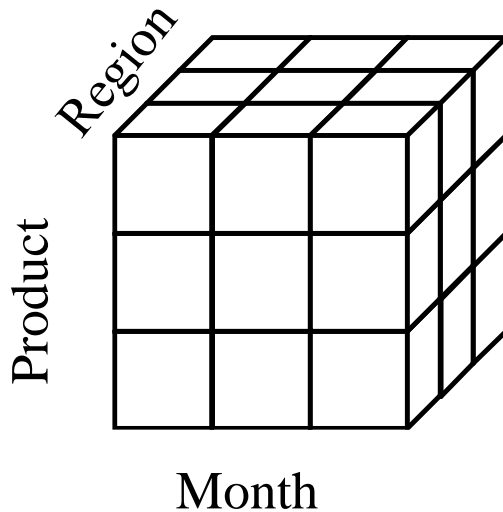
Sales volume

		Electronics	Store1	Store2
Q1		VCR	\$1,4	\$1,4
		Camcorder	\$0,6	\$0,6
		TV	\$2,0	\$2,4
		CD player	\$1,2	\$1,2
Q2		VCR	\$2,4	\$2,4
		Camcorder	\$3,3	\$1,3
		TV	\$2,2	\$2,5
		CD player	\$1,0	\$1,0

Item: Επίπεδο **Category**

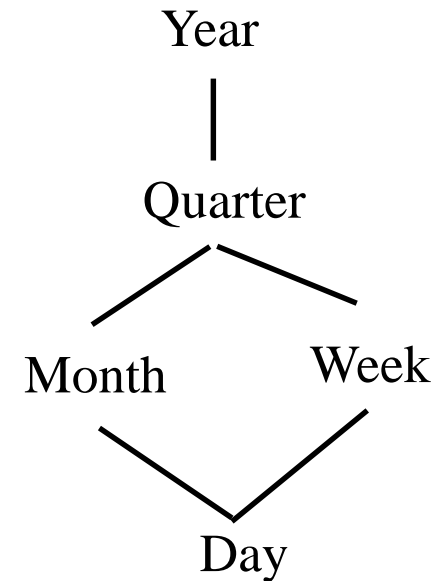
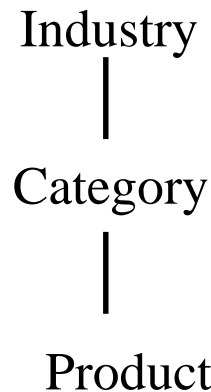
Πολυδιάστατο μοντέλο δεδομένων

Sales volume

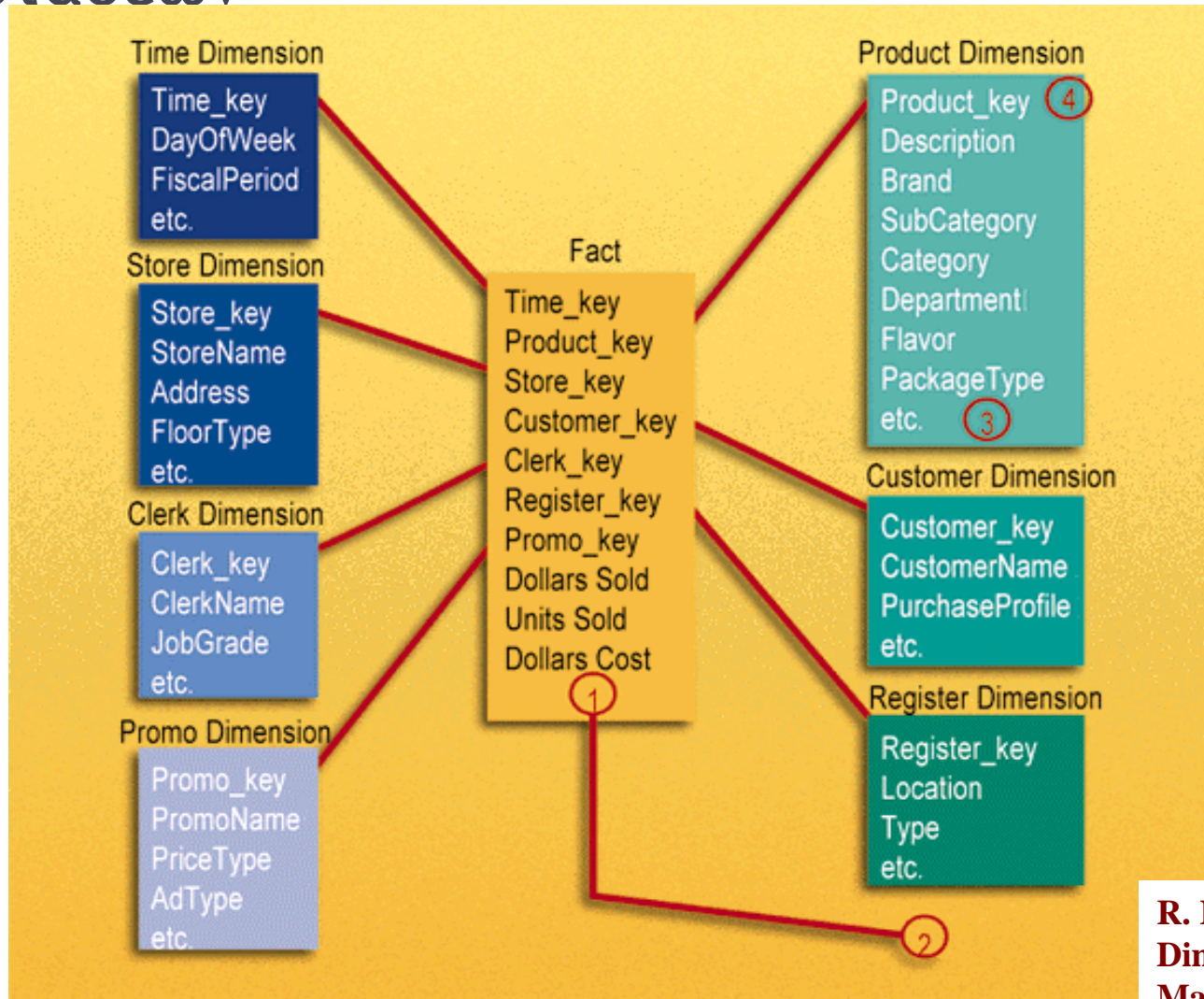


Διαστάσεις: Product, Region, Date

Ιεραρχίες διαστάσεων:



Παράδειγμα σχήματος στο Μοντέλο Διαστάσεων



**R. Kimball, A
Dimensional Modeling
Manifesto, DBMS
Magazine, Aug. 1997**

Εννοιολογική εναρμόνιση

**Source 1:
Personnel
(Cobol)**

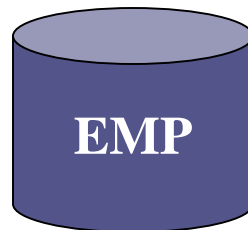
<u>EMP ID</u>	Name	DoB	Salary	Total Income	DeptID
110	Kostas	1/1/72	1500	1200	132
...

**Source 2:
Accounting
(DB2)**

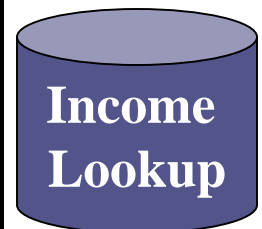
<u>EMP ID</u>	<u>IL_ID</u>	Amount
110	10	1500
110	30	300



<u>EMP ID</u>	Name	Age
110	Kostas	36
120	Nikos	48
130	Roula	29



<u>IL_ID</u>	Descr
10	Μισθός
20	Επίδομα Τέκνων
30	Φόρος
...	...

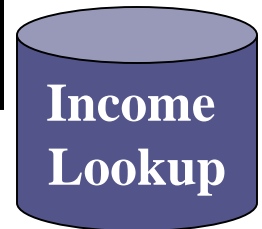


Αποκανονικοποίηση - Denormalization

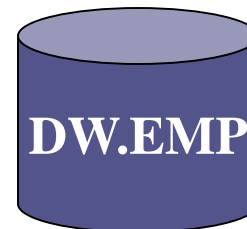
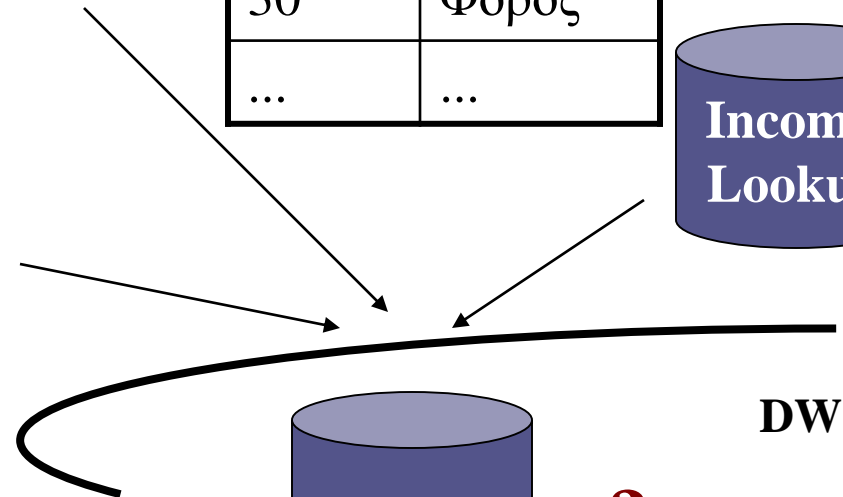
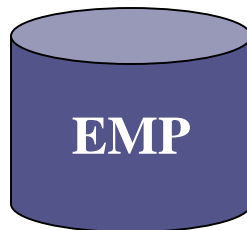
<u>EMP ID</u>	<u>IL_ID</u>	Amount
110	10	1500
110	30	300



<u>IL_ID</u>	Descr
10	Μισθός
20	Επίδομα Τέκνων
30	Φόρος
...	...



<u>EMP ID</u>	Name	Age
110	Kostas	30
120	Mitsos	48
130	Roula	29

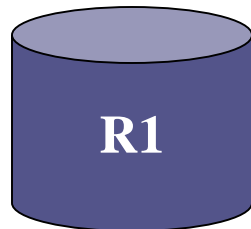


?

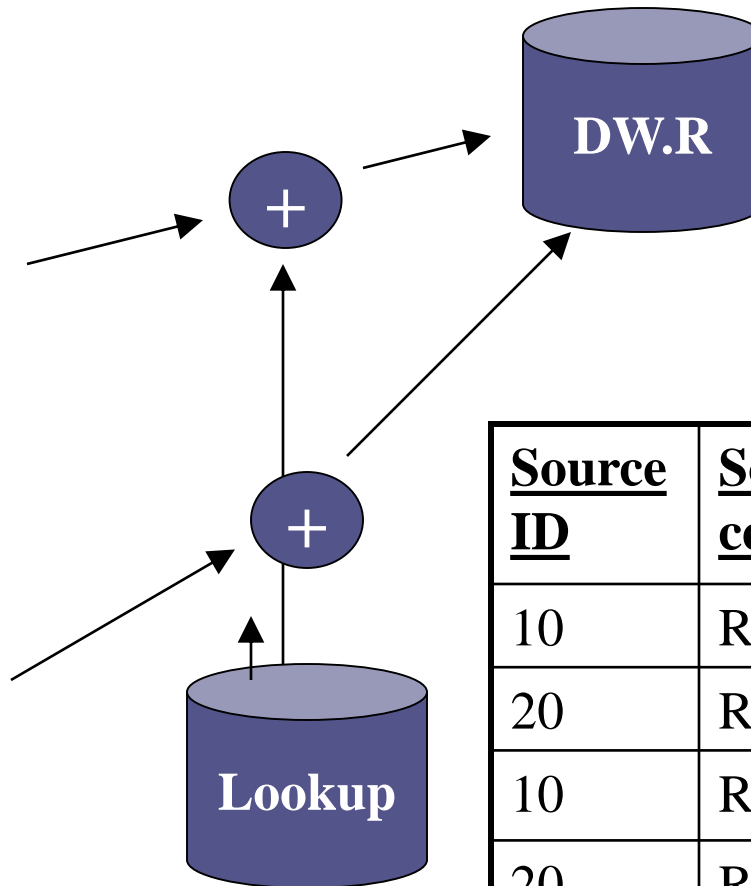
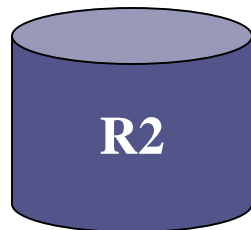
DW

Καθαρισμός των δεδομένων - Ομογενοποίηση κλειδιών

<u>ID</u>	Descr
10	Coca
20	Pepsi



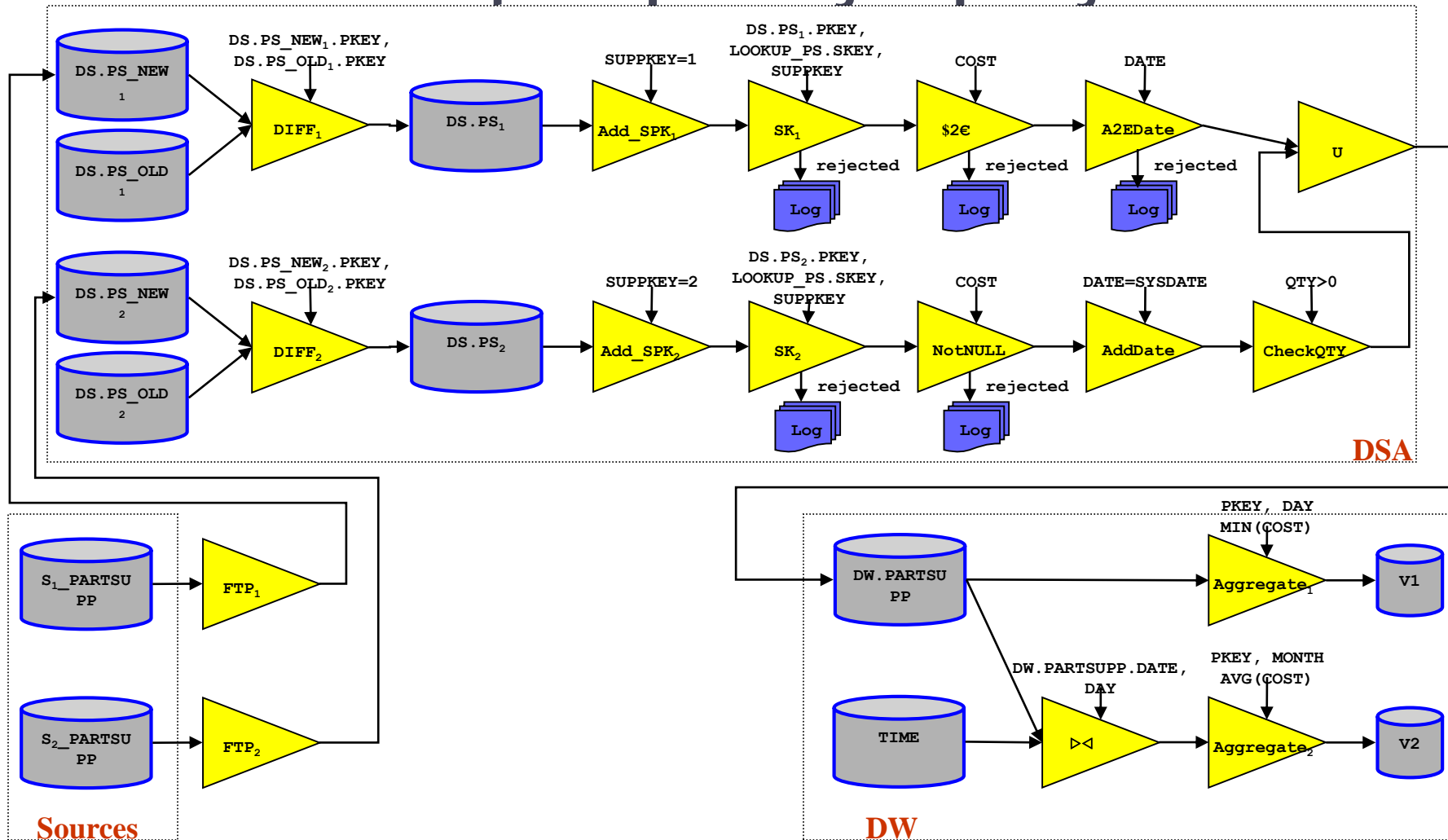
<u>ID</u>	Descr
10	Pepsi
20	HBH



<u>ID</u>	Descr
100	Coca
110	Pepsi
120	HBH

<u>Source ID</u>	<u>Sour ce</u>	Surrogate Key
10	R1	100
20	R1	110
10	R2	110
20	R2	120

DW ≠ Αποθηκευμένες όψεις!



Αντικειμενοστρεφείς ΒΔ

