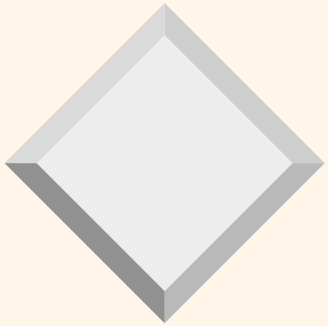


Υποστήριξη Αποφάσεων

Κεφάλαιο 23

Εισαγωγή

- Επιχειρ. αναλύουν τρέχοντα και ιστορικά δεδομ. για να ανακαλ. χρήσιμες συσχετίσεις και να υποστ. την επιχ. στρατηγική.
- Με έμφαση την πολύπλοκη, διαδραστική και διερευνητική ανάλυση (διαφορετικών κυρίως στατικών) δεδομένων που προκύπτουν από τη συνένωση πολλών ΒΔ της επιχείρησης.
  - Απευθείας Αναλυτική Επεξεργασία (OLAP) αντίθεση με Απευθείας Επεξεργασία Συναλλαγών (OLTP): Πολύπλοκα αιτήματα χωρίς συχνές ανανεώσεις δεδομένων.

Τρεις αλληλοσυμπληρώμενες τάσεις

- Αποθήκες δεδομένων: Συγκέντρωση δεδομένων από πολλές πηγές σε ενιαίο χώρο.
  - Φόρτωση, περιοδικός συγχρονισμός αντιγράφων.
  - Σημασιολογική ολοκλήρωση.
- OLAP:
  - Πολύπλοκα αιτήματα SQL και όψεις.
  - Αιτήματα βασισμένα στη φιλοσοφία των λογιστικών φύλλων και πολυδιάστατης θεώρησης των δεδομένων.
  - Διαδραστικά και άμεσα αιτήματα.
- Εξόρυξη δεδομένων: Διερευνητική αναζήτηση για τάσεις με ενδιαφέρον (Άλλο μάθημα!)

Αποθήκη Δεδομένων

- Συνηθ. Μεγέθη: gigabytes - terabytes
- Απαιτούνται χρόνοι για διαδραστικές απαντήσεις σε πολύπλοκα αιτήματα. εξειδικευμένες μεταβολές ασυνήθιστες



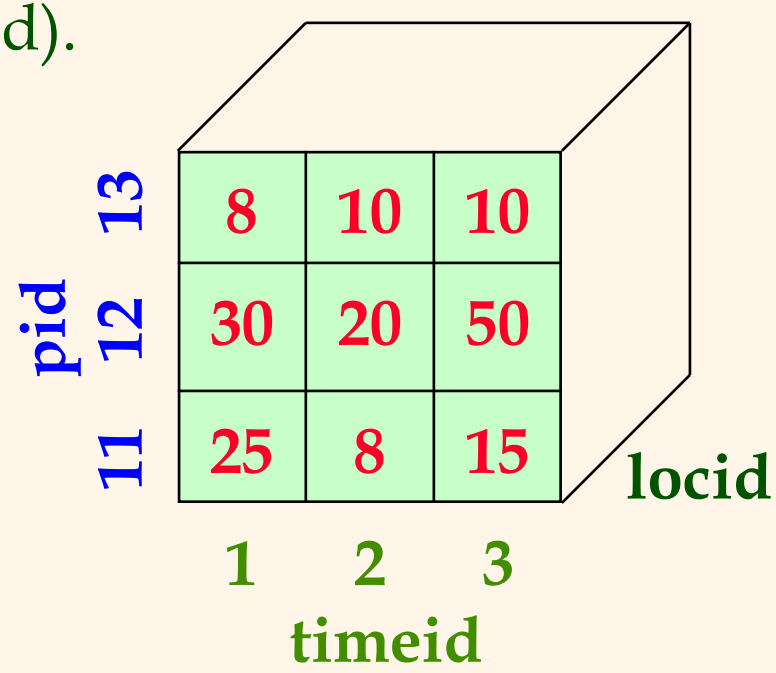
Αποθήκες δεδομένων: ζητήματα

- Σημασιολογική ολοκλήρωση: Αντιμετώπιση προβλημάτων ασυμβατότητας λόγω της ανάκτησης δεδομένων από πολλές πηγές.
- Ετερογενείς πηγές: Ανάκτηση δεδομένων από ποικίλες πηγές και αποθηκευτικού χώρους.
  - Αξιοποίηση δυνατοτήτων αντιγραφής.
- Φόρτωση, Ανανέωση, Απομάκρυνση: Φόρτωση και αναν. δεδομένων, απομάκ. πολύ παλιών.
- Διαχείριση μεταδεδομένων: Έλεγχος πηγών, χρόνου φόρτωσης και γενικών πληροφοριών για όλα τα δεδομένα στην αποθήκη.

Πολυδιάστατο μοντέλο δεδομένων

- Συλλογή αριθμητικών μετρήσεων, που εξαρτώνται από σύνολο διαστάσεων.
  - Π.χ., μέτρηση **Sales**, διαστάσεις **Product** (κλειδί: pid), **Location** (locid), και **Time** (timeid).

Παρουσίαση τομής για locid=1



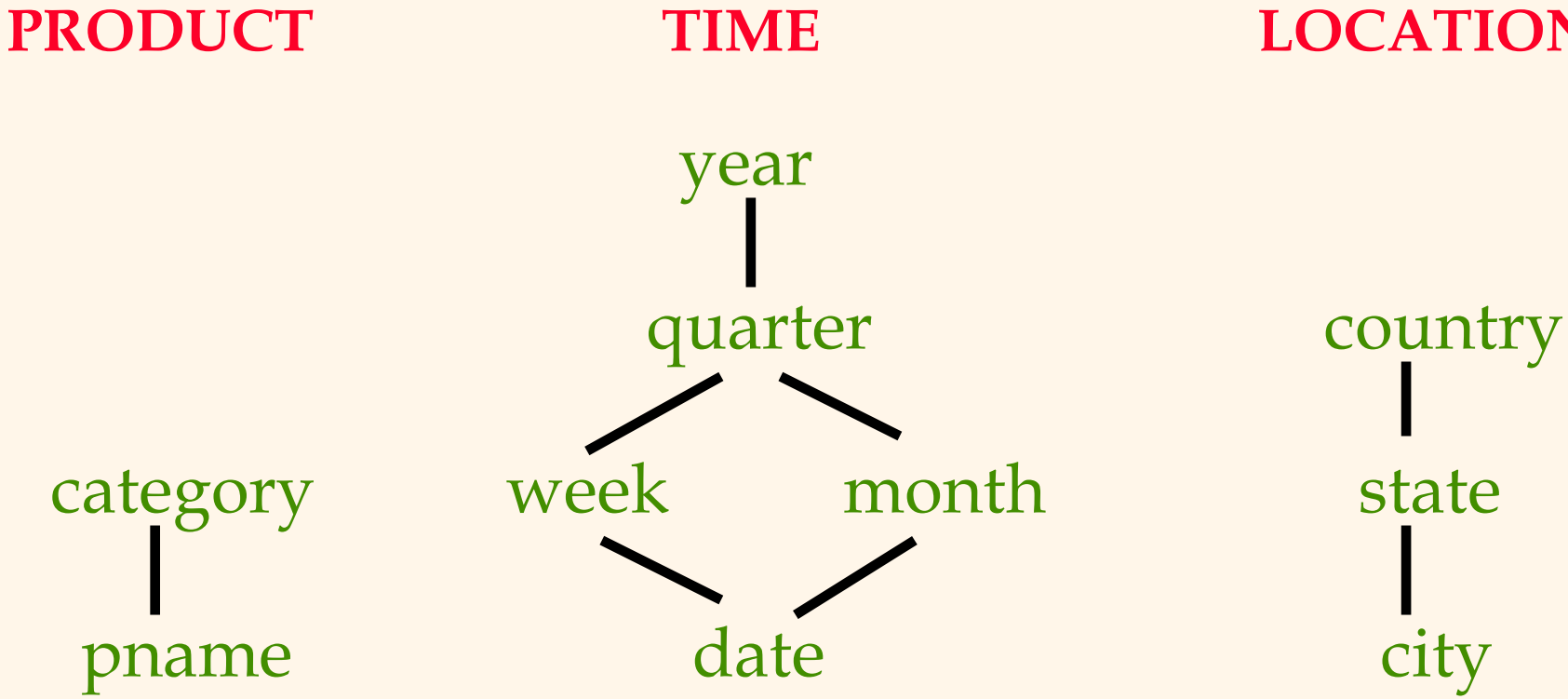
pid	timeid	locid	sales
11	1	1	25
11	2	1	8
11	3	1	15
12	1	1	30
12	2	1	20
12	3	1	50
13	1	1	8
13	2	1	10
13	3	1	10
11	1	2	35
•	•	•	

MOLAP αντ. ROLAP

- Τα πολυδιάστατα δεδομένα αποθηκεύονται είτε ως πίνακες (μόνιμα στο δίσκο): Συστήματα MOLAP. Εναλλακτικά αποθηκεύονται ως σχέσεις ROLAP συστήματα.
- Η κεντρική σχέση, που συσχετίζει τις διαστάσεις με κάποια μέτρηση ονομ. πίνακας γεγονότων. Κάθε διάσταση μπορεί να έχει επιπρ. γνωρίσματα στον πίνακα της διάστασης.
  - Π.χ., **Products(pid, pname, category, price)**
  - Οι πίνακες γεγονότων είναι μεγαλύτεροι από τους πίνακες διαστάσεων.

Ιεραρχίες Διαστάσεων

- Για κάθε διάσταση το σύνολο των τιμών οργανώνεται ιεραρχικά:



Αιτήματα OLAP (1)

- Επιρροές από SQL και λογιστικά φύλλα.
- Συνηθισμένη πράξη είναι η συνάθροιση μιας μέτρησης ως προς πλήθος διαστάσεων.
  - Σύνολο πωλήσεων.
  - Σύνολο πωλήσεων για κάθε πόλη, ή για κάθε νομό.
  - Πέντε πρώτα προϊόντα βάσει συνόλου πωλήσεων.
- Σύμπτυξη: Συνάθροιση σε διαφορετικά επίπεδα μιας ιεραρχίας διαστάσεων.
  - Π.χ., Με δεδομένο τις συνολικές πωλήσεις ανά πόλη η σύμπτυξη μπορεί να οδηγήσει στον υπολογισμό των πωλήσεων ανά νομό.

Αιτήματα OLAP (2)

- Ανάπτυξη: Το αντίθετο της σύμπτυξης.
  - Π.χ., με δεδομένο τις πωλήσεις ανά νομό, ανάπτυξη των πωλήσεων ανά πόλη.
  - Ανάπτυξη σε διαφορετική διάσταση, πχ υπολογ. συνόλου πωλήσ. ανά προϊόν για κάθε νομό.
- Περιστροφή: Συνάθροιση σε επιλεγμ. διαστάσ.
  - Π.χ., Περιστροφή σε τοποθεσία χρόνο έχει το εξής αποτέλεσμα:
- Τομές και κύβοι: Ισότητες κι επιλογές διαστήματος σε μία ή περισσότερες διαστ.

	WI	CA	Total
1995	63	81	144
1996	38	107	145
1997	75	35	110
Total	176	223	339

Σύγκριση με αιτήματα SQL

- Το αποτέλεσμα της περιστροφής μπορεί να υπολογιστεί με μια σειρά από SQL αιτήματα:

```
SELECT SUM(S.sales)
FROM   Sales S, Times T, Locations L
WHERE  S.timeid=T.timeid AND S.timeid=L.timeid
GROUP BY T.year, L.state
```

```
SELECT SUM(S.sales)
FROM   Sales S, Times T
WHERE  S.timeid=T.timeid
GROUP BY T.year
```

```
SELECT SUM(S.sales)
FROM   Sales S, Location L
WHERE  S.timeid=L.timeid
GROUP BY L.state
```

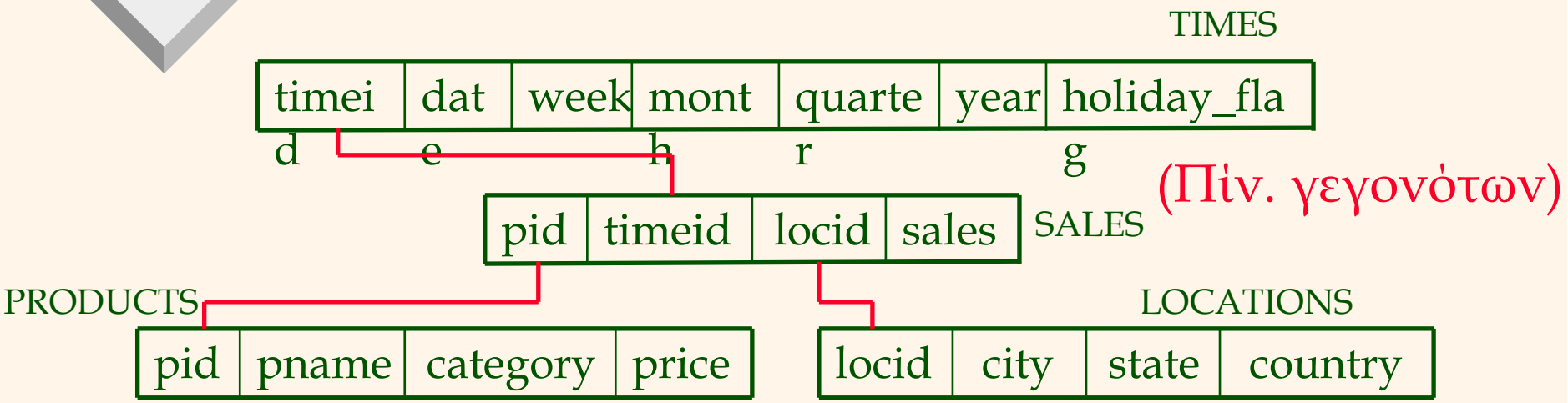
Ο τελεστής CUBE

- Γενικεύοντας το προηγούμενο παράδειγμα, αν υπάρχουν k διαστάσεις, έχουμε 2^k πιθανά SQL αιτήματα τύπου GROUP BY που προκύπτουν από την περιστροφή υποσυνόλου διαστάσεων.
- CUBE pid, locid, timeid BY SUM Sales
  - Ισοδύναμο με τη σύμπτυξη των πωλήσεων στα οκτώ υποσύνολα του συνόλου {pid, locid, timeid}. Κάθε σύμπτυξη αντιστοιχεί σε ένα αίτημα της μορφής:

Αρκετή έρευνα σχετικά με τη βελτιστοποίηση του τελεστή CUBE !

```
SELECT SUM(S.sales)
FROM   Sales S
GROUP BY grouping-list
```

Ζητήματα σχεδιασμού



- Πίνακας γεγονότων σε BCNF. Πίνακες διαστάσεων όχι κανονικοποιημένοι.
  - Πίνακες διαστάσεων μικροί. Σπάνιες μεταβολές. Σημαντική η γρήγορη επεξεργασία αιτημάτων.
- Το σχήμα αυτό είναι συν. σε εφαρμογές OLAP. Ονομάζεται σχήμα άστρου. Ο υπολογισμός της σύζευξης όλων των πινάκων: σύζευξη άστρου.

Περίληψη

- Η υποστήριξη αποφάσεων είναι κομμάτι των βάσεων δεδομένων.
- Προϋποθέτει τη δημιουργία αποθηκών δεδομένων.
- Οι αποθήκες δεδομένων αξιοποιούνται χρησιμοποιώντας τεχνικές ανάλυσης: πολύπλοκα SQL αιτήματα και πολυδιάστατα αιτήματα OLAP (επιρροές από SQL και λογιστικά φύλλα).