

Κεφάλαιο 1

- Εισαγωγή και Ορισμοί σε Βάσεις Δεδομένων
- Η Ύλη του μαθήματος συνοπτικά

Βάσεις Δεδομένων – Περιεχόμενα Μαθήματος (1)

- **Σύντομη Εισαγωγή και Ιστορική Αναδρομή**
 - Αρχιτεκτονικές Συστημάτων
- **Μοντελοποίηση, Σχεδιασμός και Ανάπτυξη Β.Δ.**
 - Το Μοντέλο Entity-Relationship (E-R)
 - Άλλα εννοιολογικά Μοντέλα (UML)
- **Το Σχεσιακό (Relational) Μοντέλο**
 - Ορισμοί
 - Σχεσιακή Άλγεβρα και Σχεσιακός Λογισμός
- **Σχεσιακά Συστήματα – η Γλώσσα SQL**
 - Η SQL και άλλες σχεσιακές γλώσσες
 - Ακεραιότητα, Δικαιοδοσία και Ασφάλεια
 - Θεωρία Σχεδιασμού Βάσεων Δεδομένων – Κανονικοποίηση

Βάσεις Δεδομένων – Περιεχόμενα Μαθήματος (2)

■ Φυσική Οργάνωση Βάσεων Δεδομένων

- Δομές Αποθήκευσης και Συστήματα Αρχείων
- Ευρετήρια και Hash
- Επεξεργασία και Βελτιστοποίηση Επερωτήσεων / Ερωτημάτων
- Βελτιστοποίηση Ερωτημάτων

■ Αρχιτεκτονικές, Συστήματα και Λειτουργικά Θέματα

- **Relational:** Oracle, IBM DB2, SQL Server, MySQL,
- Νέες / Αναδυόμενες Τεχνολογίες και Συστήματα
 - XML – based:** BaseX, Sedna, eXist, ... Ανοικτός Κώδικας, XQUERY
 - Object-Oriented DBMS:** ObjectivityDB, Versant, ...
 - Object-Relational:** Hibernate (Java), ActiveRecord (Ruby), ...
 - NoSQL Database Systems:** CouchDB, MongoDB, BigTable, Hbase, Apache Cassandra, Hypertable...

Εισαγωγή - Ορισμοί

- ***ΔΕΔΟΜΕΝΑ (DATA)***

Γνωστά γεγονότα / περιστατικά που καταγράφονται και έχουν νόημα

- ***ΒΑΣΗ ΔΕΔΟΜΕΝΩΝ (DATABASE)***

Μία ΟΛΟΚΛΗΡΩΜΕΝΗ συλλογή Δεδομένων που συσχετίζονται, η οποία αποθηκεύεται (συνήθως) σε δευτερεύουσα Μνήμη

Η Βάση Δεδομένων αναπαριστά ένα υποσύνολο των στοιχείων που μπορούμε να συλλέξουμε για τον μικρόκοσμο(πεδίο αναφοράς) μιας εφαρμογής.

- ***ΣΥΣΤΗΜΑ ΔΙΟΙΚΗΣΗΣ / ΔΙΑΧΕΙΡΙΣΗΣ ΒΑΣΕΩΝ ΔΕΔΟΜΕΝΩΝ - ΣΔΒΔ (DATABASE MANAGEMENT SYSTEM - DBMS)***

Μια αυτοτελής συλλογή από τμήματα λογισμικού (προγράμματα) για την δημιουργία, επεξεργασία και την συντήρηση βάσεων δεδομένων.

- ***ΣΥΣΤΗΜΑ ΒΑΣΕΩΝ ΔΕΔΟΜΕΝΩΝ (DATABASE SYSTEM)***

Το λογισμικό (DBMS) μαζί με τη Βάση Δεδομένων.

Database Management Systems

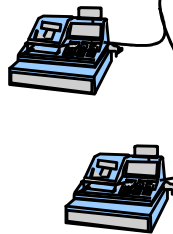


Αναφορές και
Ερωτήματα

Database

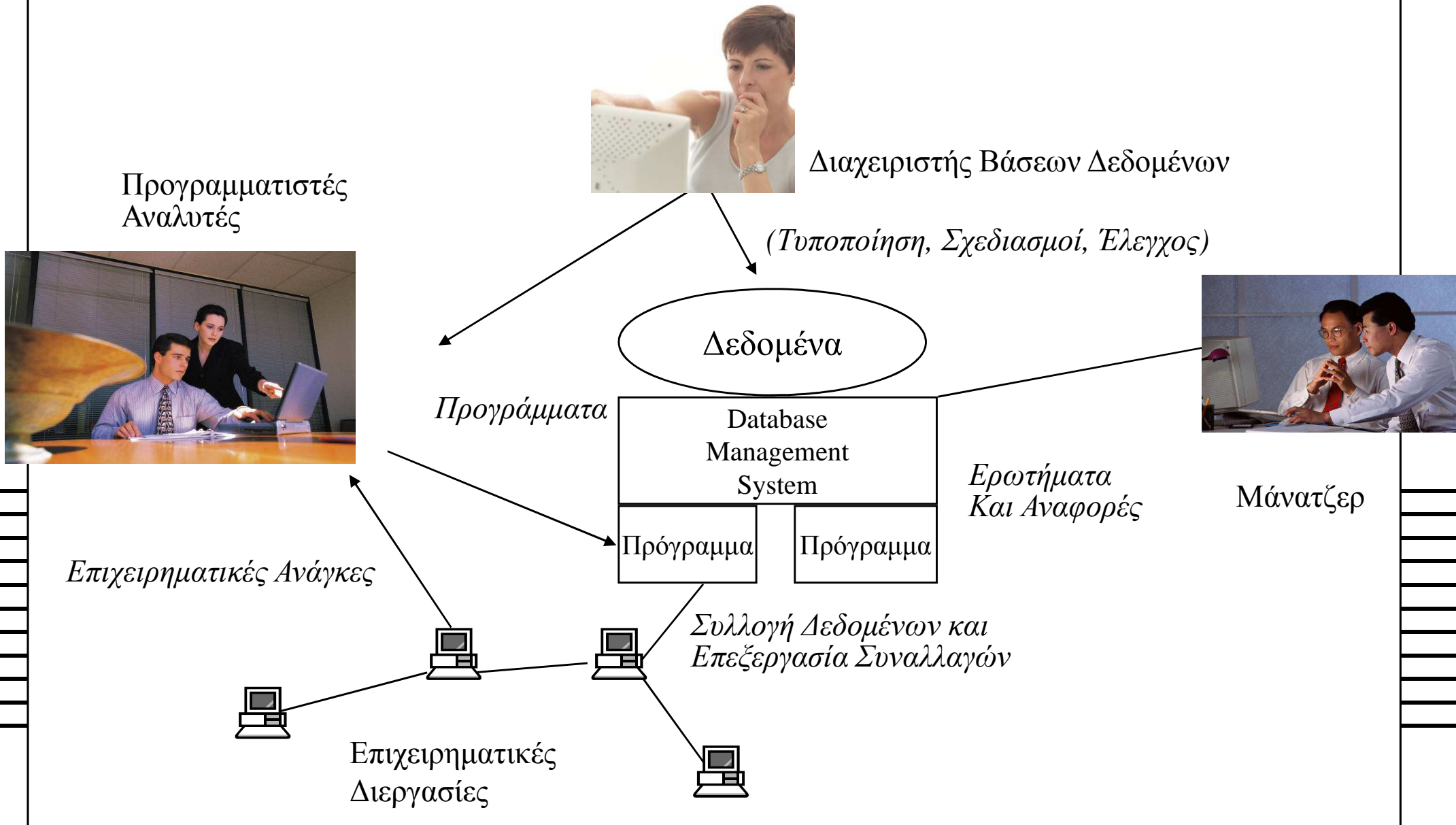
DBMS

Programs



Πωλήσεις και Δεδομένα Συναλλαγών
(transactions)

Βασικοί Ρόλοι με ένα DBMS



Τι είναι ένα ΣΔΒΔ?

1. Διαχειρίζεται μεγάλο όγκο δεδομένων.
2. Υποστηρίζει αποτελεσματική πρόσβαση σε μεγάλο όγκο δεδομένων
3. Υποστηρίζει ταυτόχρονη πρόσβαση πολλών χρηστών σε μεγάλο όγκο δεδομένων.
 - Παράδειγμα: Αεροπορική Εταιρεία και Πράκτορες.
4. Υποστηρίζει ασφαλή, ατομική πρόσβαση σε μεγάλο όγκο δεδομένων
 - Αν δύο άνθρωποι «τραβούν» χρήματα την ίδια χρονική στιγμή από τον ίδιο λογαριασμό (μέσω ΑΤΜ) – το υπόλοιπο του λογαριασμού θα είναι ΛΑΘΟΣ όποιος και αν ολοκληρώσει τη συναλλαγή τελευταίος.

3 ΟΥΣΙΕΣ ΤΩΝ ΒΑΣΕΩΝ ΔΕΔΟΜΕΝΩΝ

1. Μοντελοποίηση και Σχεδιασμός ΒΔ
 - Τι κάνουμε πριν από την υλοποίηση και λειτουργία ΒΔ.
2. Προγραμματισμός: Ερωτήσεις και Πράξεις σε ΒΔ
 - SQL = “παγκόσμιος γλώσσα”
3. Υλοποίηση Συστημάτων ΒΔ.

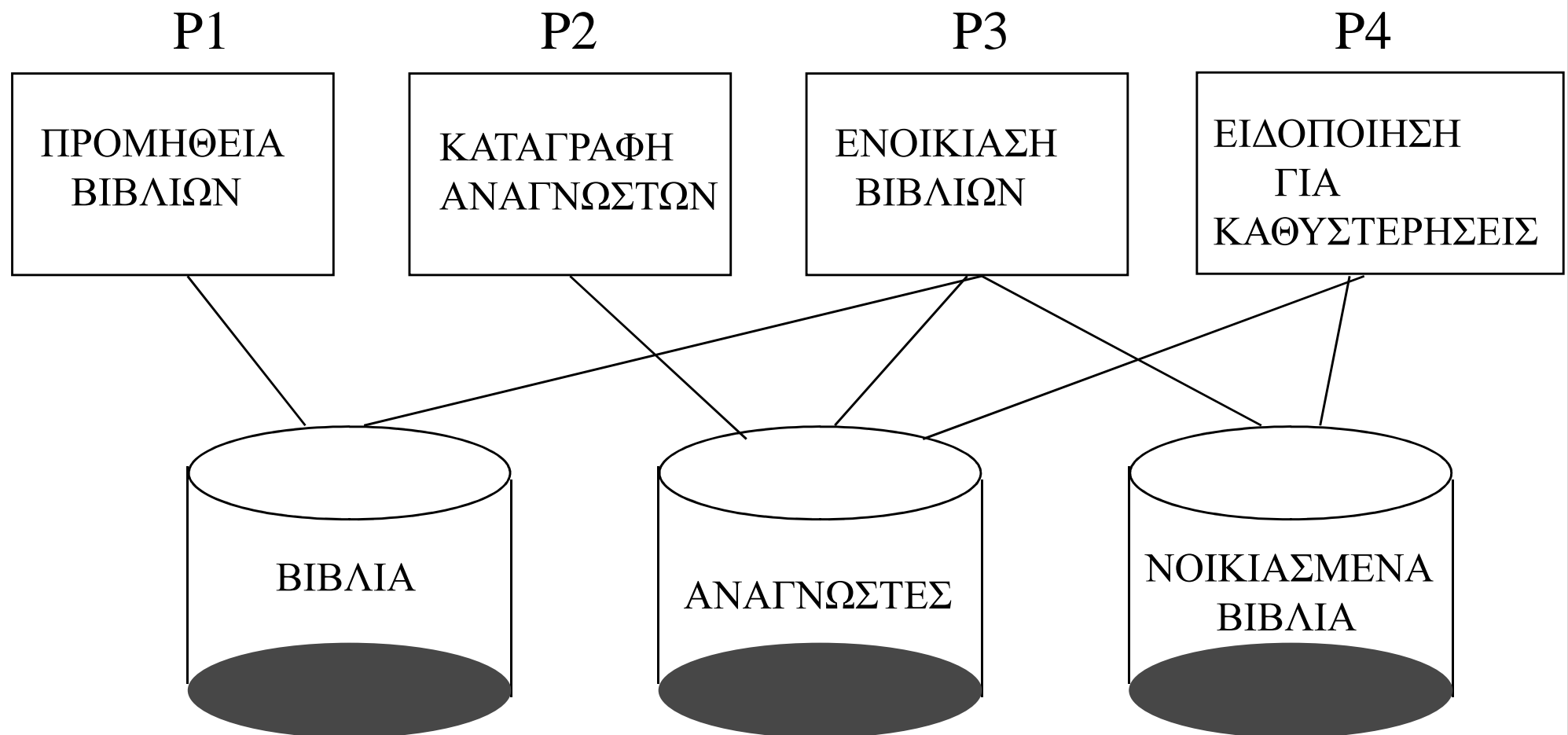
7^ο Εξάμηνο = (1) + (2),

9^ο Εξάμηνο = (3)

Ιστορική Αναδρομή (1)

- 1950s (Πρώτη Γενεά - Αρχεία σε Ταινίες)
 - batch processing, κάρτες και ταινίες (σειριακή επεξεργασία)
- 1960s (Δεύτερη Γενεά - Αρχεία σε Δίσκους)
 - Εκτεταμένη χρήση των Δίσκων
 - » Επιστημονικό Πεδίο Βάσεων Δεδομένων
 - Πρώτα Συστήματα Αρχείων
 - Γενικευμένα Πακέτα Ταξινόμησης Δεδομένων
 - Δειλή παρουσία Γενικευμένων Λογισμικών Πακέτων
 - Ορισμός των Δεδομένων σε Γλώσσα Προγραμματισμού
 - » **COBOL**
 - Ανάπτυξη Συστημάτων ΒΔ για Ιδιωτική Χρήση

Το Σενάριο για Συστήματα Αρχείου



ΕΦΑΡΜΟΓΕΣ πάνω σε Συστήματα Αρχείων

- Μειονεκτήματα στη χρήση Αρχείων για αποθήκευση δεδομένων:
 - Πλεονασμός / Επαναληπτικότητα Δεδομένων και Ασυνέπειες
 - » Πολλαπλές Μορφές, Επανάληψη των Δεδομένων σε αρχεία
 - Δυσκολίες πρόσβασης στα δεδομένα
 - » Για κάθε νέα ανάγκη πρέπει να γραφεί νέο (μεγάλο) πρόγραμμα
 - Απομόνωση των δεδομένων σε πολλαπλά αρχεία διαφορετικής μορφής → Δυσκολίες για υλοποίηση νέων εφαρμογών
 - Προβλήματα Ακεραιότητας
 - » Οι κανόνες ακεραιότητας (π.χ., υπόλοιπο λογαριασμού > 0) πρέπει να γίνουν μέρος του κώδικα
 - » Δύσκολο να εισαχθούν νέοι κανόνες ή να αλλάξουν οι παλιοί

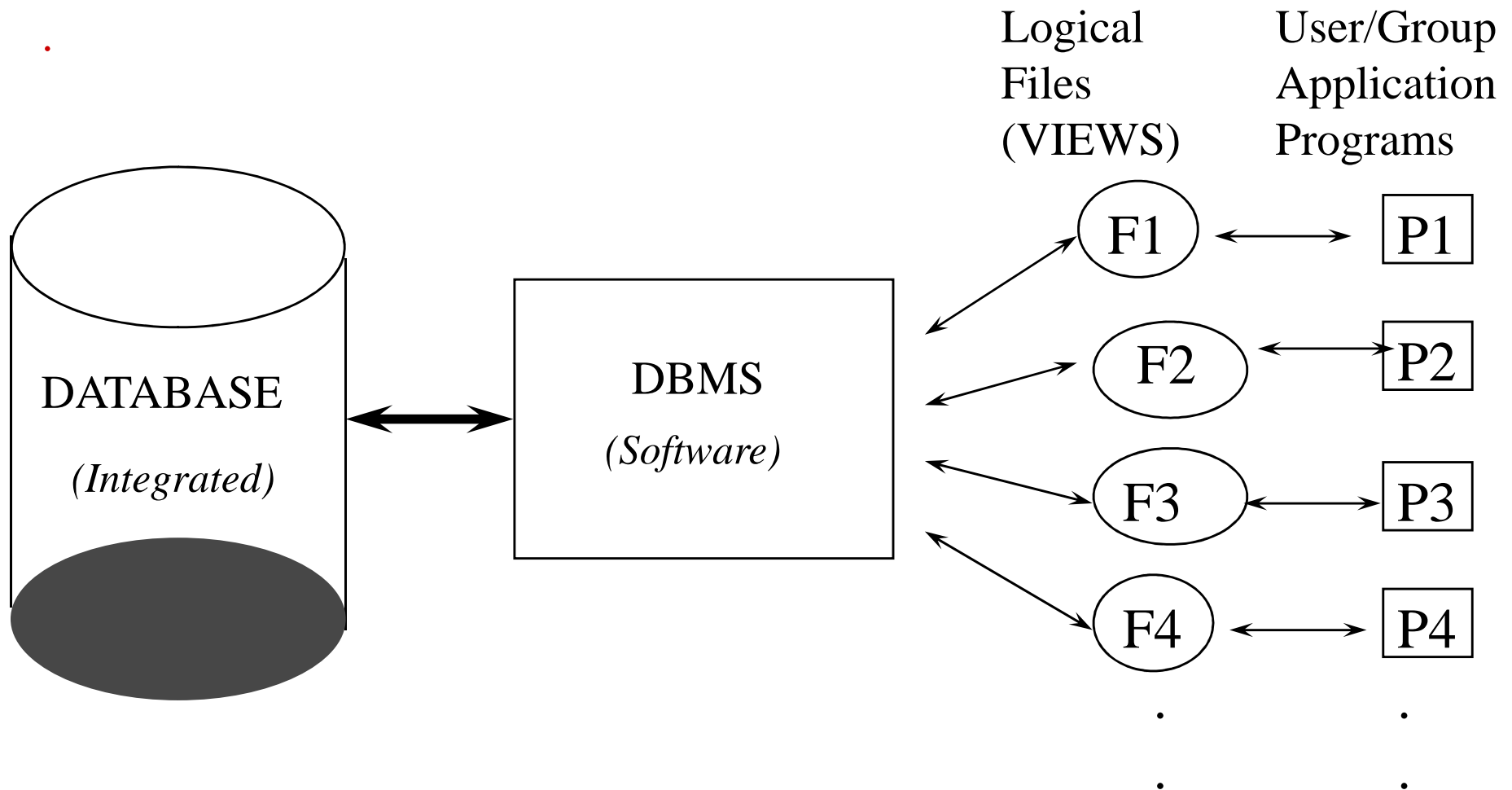
ΕΦΑΡΜΟΓΕΣ πάνω σε Συστήματα Αρχείων

- Μειονεκτήματα στη χρήση Αρχείων για αποθήκευση δεδομένων (συνέχεια):
 - Ατομικότητα των Αλλαγών
 - » Σε περίπτωση αστοχίας του συστήματος, υπάρχει κίνδυνος να μείνει η βάση δεδομένων σε ενδιάμεση κατάσταση (δηλαδή, με λάθη)
 - » Ενδεικτικά, η μεταφορά χρημάτων από λογαριασμό σε λογαριασμό ή πρέπει να γίνεται πλήρως ή καθόλου
 - Ταυτόχρονη Χρήση από πολλούς χρήστες
 - » Απαιτείται για καλύτερη απόδοση
 - » Όταν δεν υπάρχει έλεγχος στην ταυτόχρονη χρήση φτάνουμε σε λάθη
 - Π.χ., δύο χρήστες διαβάζουν το υπόλοιπο και το αλλάζουν ταυτόχρονα
 - Προβλήματα Ασφαλείας
 - » Ποιος έχει πρόσβαση σε ποιο αρχείο?
- Τα ΣΔΒΔ προσφέρουν ΛΥΣΕΙΣ ΣΕ ΟΛΑ ΤΑ ΠΑΡΑΠΑΝΩ

Συστήματα Διαχείρισης / Διοίκησης Βάσεων Δεδομένων

- Τι θέλουμε παραπάνω από ότι έχουμε σε ένα Σύστημα Αρχείων?
 - **Απλές, αποδοτικές** επερωτήσεις (ερωτήσεις) που μπορεί να γίνονται ανάλογα με το θέμα / πρόβλημα που προκύπτει στη χρήση των δεδομένων
 - Έλεγχο Συνδρομικότητας / συγχρονισμού
 - Ανάκαμψη από κάθε πρόβλημα
 - Όλα τα καλά από τη δυνατότητα Μοντελοποίησης

Το Σενάριο Επεξεργασίας με DBMS



Ιστορική Αναδρομή (2)

■ 1970s Τρίτη Γενιά – Προ-Σχεσιακή (Pre-Relational)

- Τάση για Πρότυπα με την CODASYL αναφορές προτύπων 1969, 19, 73, 78, 81, 85...
 - » DBTG (Data Base Task Group)
- ΑΠΟΘΗΚΕΥΣΗ δεδομένων ΚΑΙ ΤΩΝ ΟΡΙΣΜΩΝ τους
- εμφύτευση γενικής χρήσης προγραμμάτων προσπέλασης σε HOST γλώσσα (COBOL)
- Συστήματα Δικτύου (NETWORK) και Ιεραρχικά (HIERARCHICAL)
- Θεωρητικός Ορισμός του Σχεσιακού Μοντέλου (από τον CODD)
- Ενδιαφέρον από την Ακαδημαϊκή / Ερευνητική Κοινότητα (Computer Science)
- Καθαρός Διαχωρισμός μεταξύ «φυσικής» και «λογικής» οργάνωσης ΒΔ
- Θεωρητικές προσεγγίσεις σε Πρακτικά Θέματα Λειτουργίας
- Τα πρώτα πιλοτικά Σχεσιακά Συστήματα (SYSTEM-R, INGRESS)
- Αναγνώριση του ρόλου της Μοντελοποίησης.
- ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ ΣΥΣΤΗΜΑΤΩΝ
 - Επεξεργασία ΒΔ σε ΧΑΜΗΛΟ επίπεδο (κακό στην παραγωγικότητα, λάθη, κλπ.)

Παράδειγμα Βάσης Δεδομένων

■ Ένας Μικρόκοσμος - τμήμα των πληροφοριών για την λειτουργία ενός Καταστήματος

Μερικές ΟΝΤΟΤΗΤΕΣ του Μικρόκοσμου (ενδιαφέρουσες)

Employees (Υπάλληλοι)

Products (Προϊόντα)

Suppliers (Προμηθευτές)

Orders (Παραγγελίες)

Customers (Πελάτες)

Μερικές ΣΥΣΧΕΤΙΣΕΙΣ μεταξύ Οντοτήτων

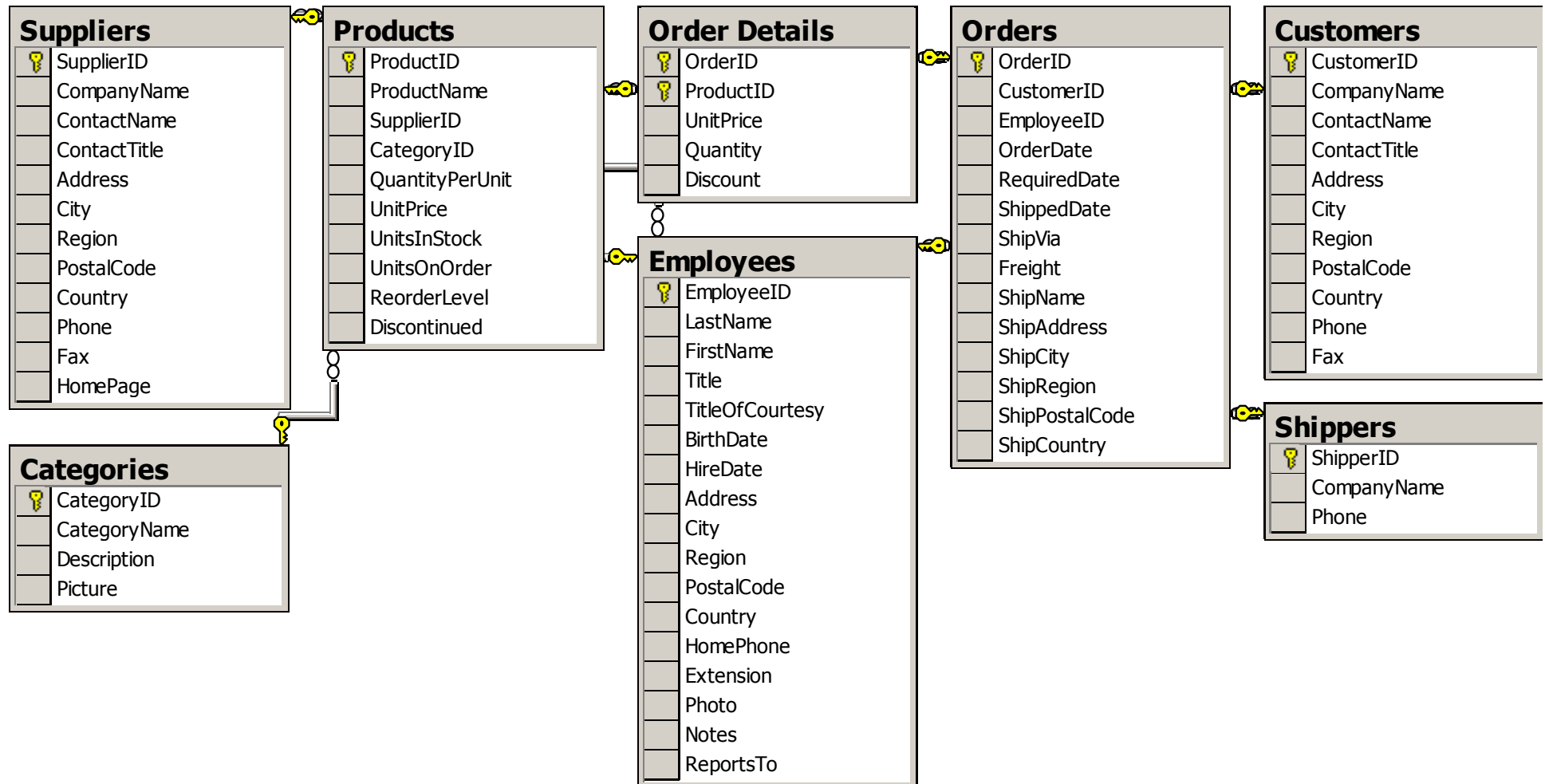
Οι Υπάλληλοι ετοιμάζουν Παραγγελίες

Οι Προμηθευτές στέλνουν Προϊόντα

Τα Προϊόντα αποστέλλονται στους Πελάτες

Οι Πελάτες παραγγέλλουν Προϊόντα

Παράδειγμα Βάσης Δεδομένων



Παράδειγμα Βάσης - Στιγμιότυπο (1)

EmployeeID	LastName	FirstName	Title	BirthDate	HireDate	ReportsTo
8	Callahan	Laura	Inside Sales Coordinator	9/1/1958	5/3/1994	2
5	Buchanan	Steven	Sales Manager	4/3/1955	17/10/1993	2
1	Davolio	Nancy	Sales Representative	8/12/1968	1/5/1992	2
3	Leverling	Janet	Sales Representative	30/8/1963	1/4/1992	2
4	Peacock	Margaret	Sales Representative	19/9/1958	3/5/1993	2
6	Suyama	Michael	Sales Representative	2/7/1963	17/10/1993	5
7	King	Robert	Sales Representative	29/5/1960	2/1/1994	5
9	Dodsworth	Anne	Sales Representative	2/7/1969	15/11/1994	5
2	Fuller	Andrew	Vice President, Sales	19/2/1952	14/8/1992	

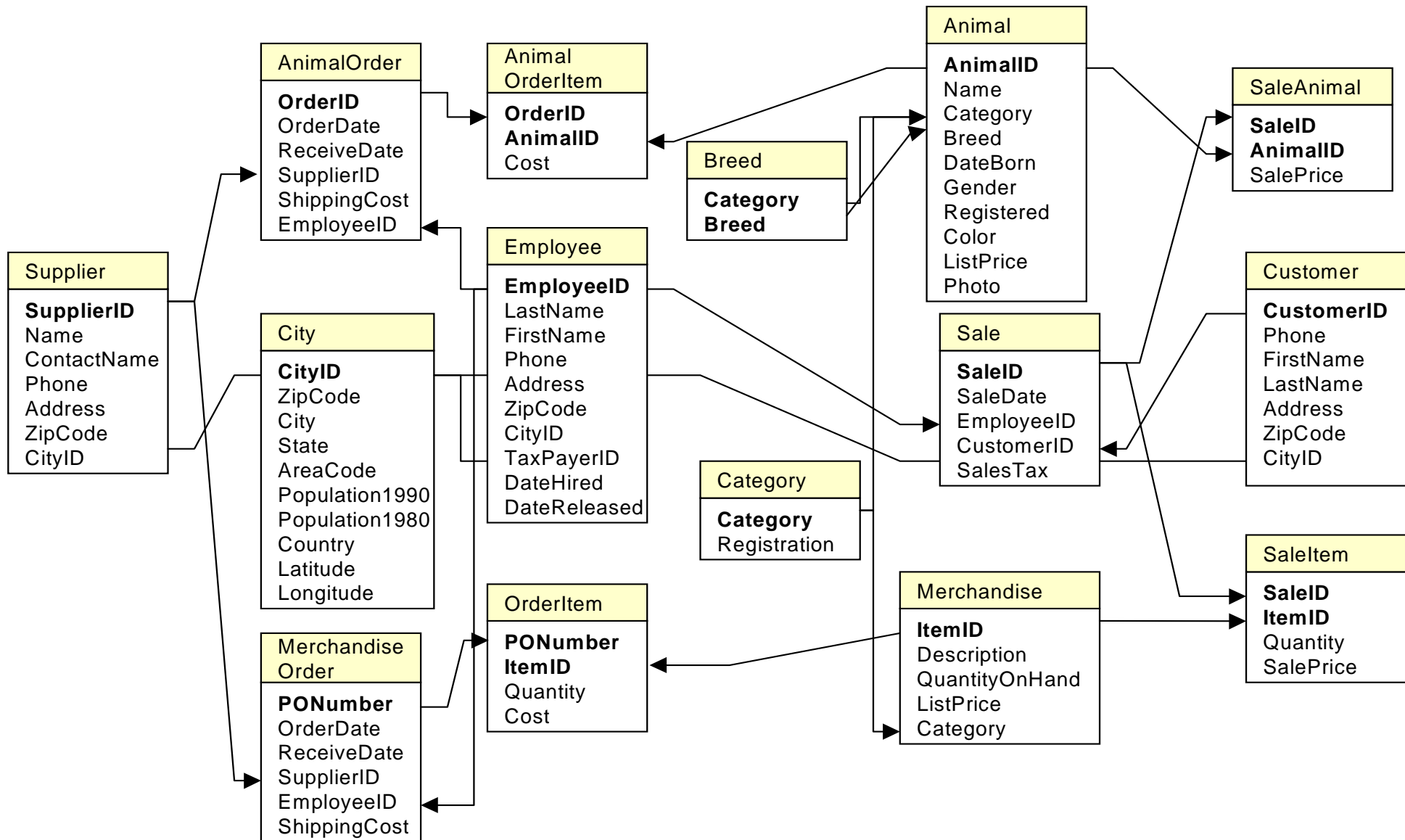
EMPLOYEES

Παράδειγμα Βάσης - Στιγμιότυπο (2)

Products

ProductID	ProductName	QuantityPerUnit	UnitPrice	UnitsInStock	CategoryName
1	Chai	10 boxes x 20 bags	18,00 Δρχ	39	Beverages
2	Chang	24 - 12 oz bottles	19,00 Δρχ	17	Beverages
3	Aniseed Syrup	12 - 550 ml bottles	10,00 Δρχ	13	Condiments
4	Chef Anton's Cajun Seasoning	48 - 6 oz jars	22,00 Δρχ	53	Condiments
6	Grandma's Boysenberry Spread	12 - 8 oz jars	25,00 Δρχ	120	Condiments
7	Uncle Bob's Organic Dried Pears	12 - 1 lb pkgs.	30,00 Δρχ	15	Produce
8	Northwoods Cranberry Sauce	12 - 12 oz jars	40,00 Δρχ	6	Condiments
10	Ikura	12 - 200 ml jars	31,00 Δρχ	31	Seafood

Περιγραφή μιας πιο ρεαλιστικής Βάσης



Ιστορική Αναδρομή (3)

■ 1980-1990s (Τετάρτη Γενεά – Σχεσιακό Μοντέλο και Συστήματα)

» Relational Database Systems

- ΥΨΗΛΟΥ επιπέδου, ισχυρότατες Γλώσσες Επεξεργασίας
 - Set-Oriented (όχι Record-oriented)
- Αναγνώριση της Θεωρίας Βάσεων Δεδομένων
- Αρχιτεκτονικές 3 επιπέδων
- Το DBMS ενσωματωμένο σε μεγάλα Συστήματα Δοσοληψιών (**Transactional Systems**) π.χ., σε Δίκτυα.
- SQL
- Εμφάνιση των Αντικειμενοστραφών Συστημάτων, Ευφυών Συστημάτων, κλπ.

Σύνοψη Εξέλιξης των ΣΔΒΔ

60' s

Ιεραρχικά

Δικτύου DBTG

70's

80's

Σχεσιακά

Η Επιλογή στις περισσότερες εφαρμογές

90' s

Αντικειμενοστρεφή

Γνωσιακά

Το Σχεσιακό Μοντέλο σε μια Διαφάνεια

.Ο Μικρόκοσμος των Υπαλλήλων – Τμημάτων όπου εργάζονται

Department	<i>dno</i>	<i>name</i>
	10	Toy
	20	Shoe

Employee	<i>eno</i>	<i>name</i>	<i>salary</i>	<i>dept</i>
	1	Lou	10000000	10
	7	Laura	150000	20
	22	Mike	80000	20

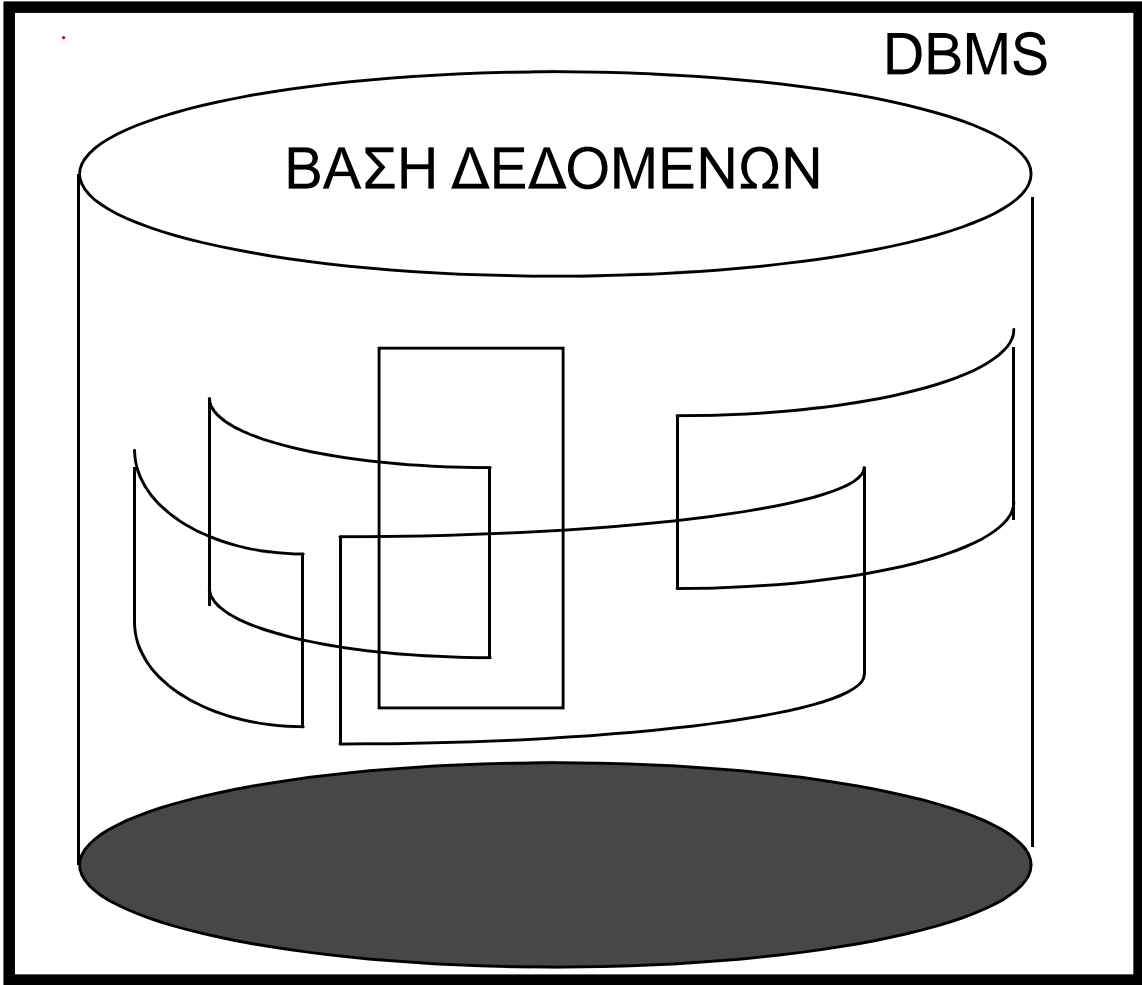


```
select E.name, E.salary, D.no  
from Employee E, Department D  
where E.salary < 100000  
      and D.name = 'Shoe'  
      and E.dept = D.dno;
```

Χαρακτηριστικά και Ιδιότητες των DBMS

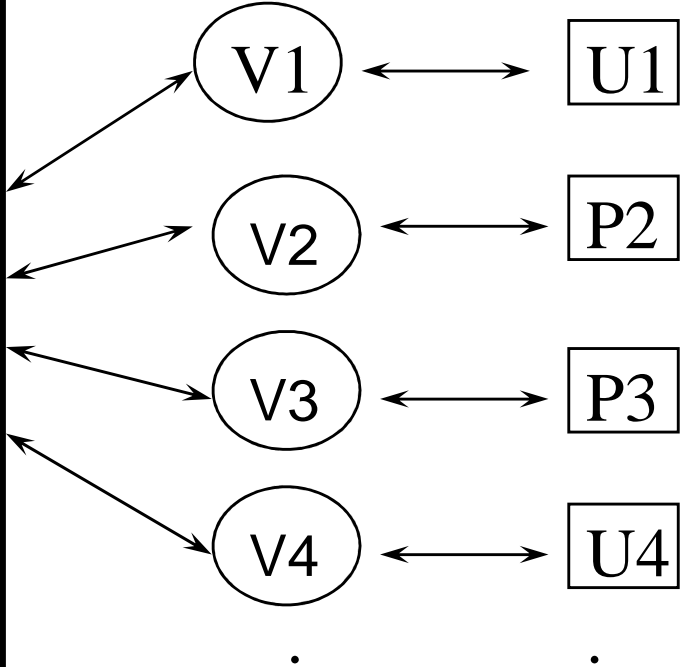
- **Αυτό-περιγραφική Φύση των Συστημάτων Βάσεων Δεδομένων – Ομοιογένεια στην Δημιουργία, Πρόσβαση και Έλεγχο των Δεδομένων**
- **Μοίρασμα (SHARING) δεδομένων και χρήση από πολλούς χρήστες**
- **Υποστήριξη κανόνων Ακεραιότητας, Ασφάλειας, Δικαιοδοσίας στα Δεδομένα**
- **Κεντρικός έλεγχος Λειτουργίας (Database Administration)**
- **Πολλαπλοί Τρόποι Πρόσβασης**
- **Έλεγχος Συνδρομικότητας και Μηχανισμοί για Ανάκαμψη**
- **Τυποποιήσεις**
- **Αύξηση της Παραγωγικότητας στην Ανάπτυξη Εφαρμογών**
- **Απομόνωση Προγραμμάτων και Δεδομένων – ΑΦΑΙΡΕΣΗ ΔΕΔΟΜΕΝΩΝ (Abstraction)**
- **ΑΠΟΔΟΣΗ !!! (Performance - Performance – Performance)**

Απλή Εικόνα ενός DBMS

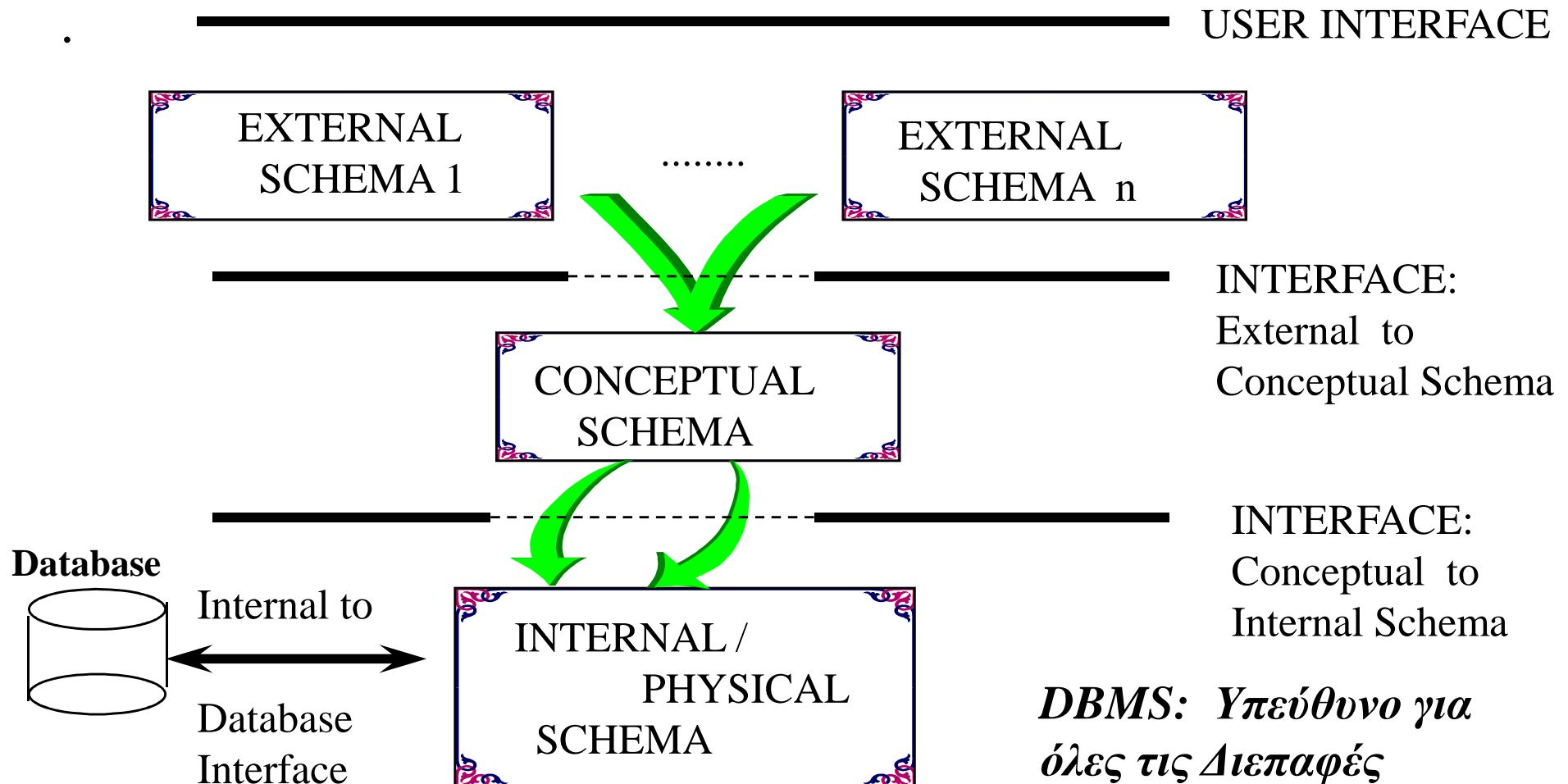


ΟΨΕΙΣ της
ΒΑΣΗΣ

ΧΡΗΣΤΕΣ ή
Προγράμματα
Εφαρμογών

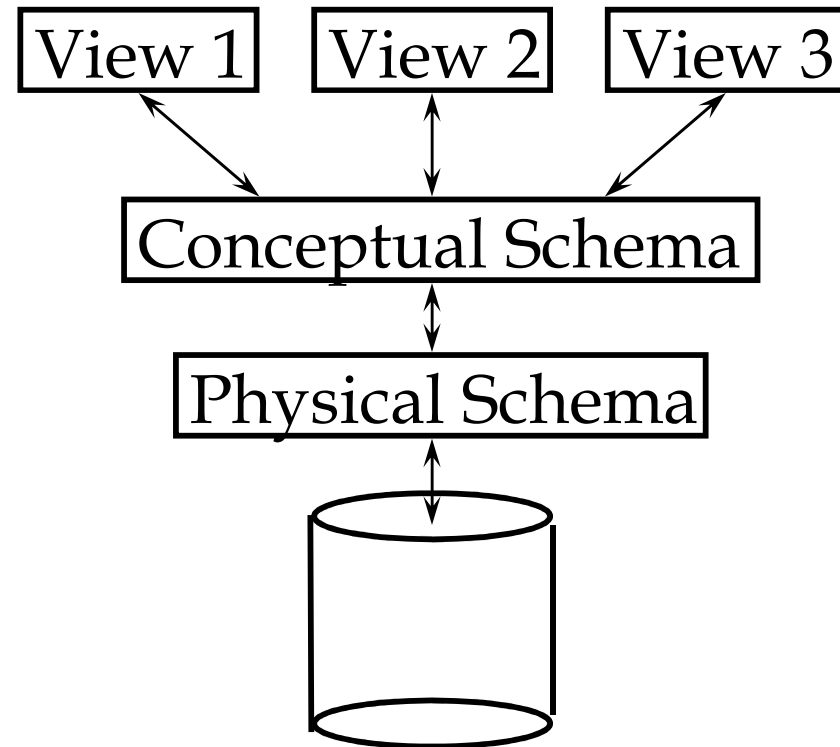


Η τυποποιημένη Αρχιτεκτονική 3-επιπέδων (ANSI/SPARC)



Αφαιρετικά Επίπεδα

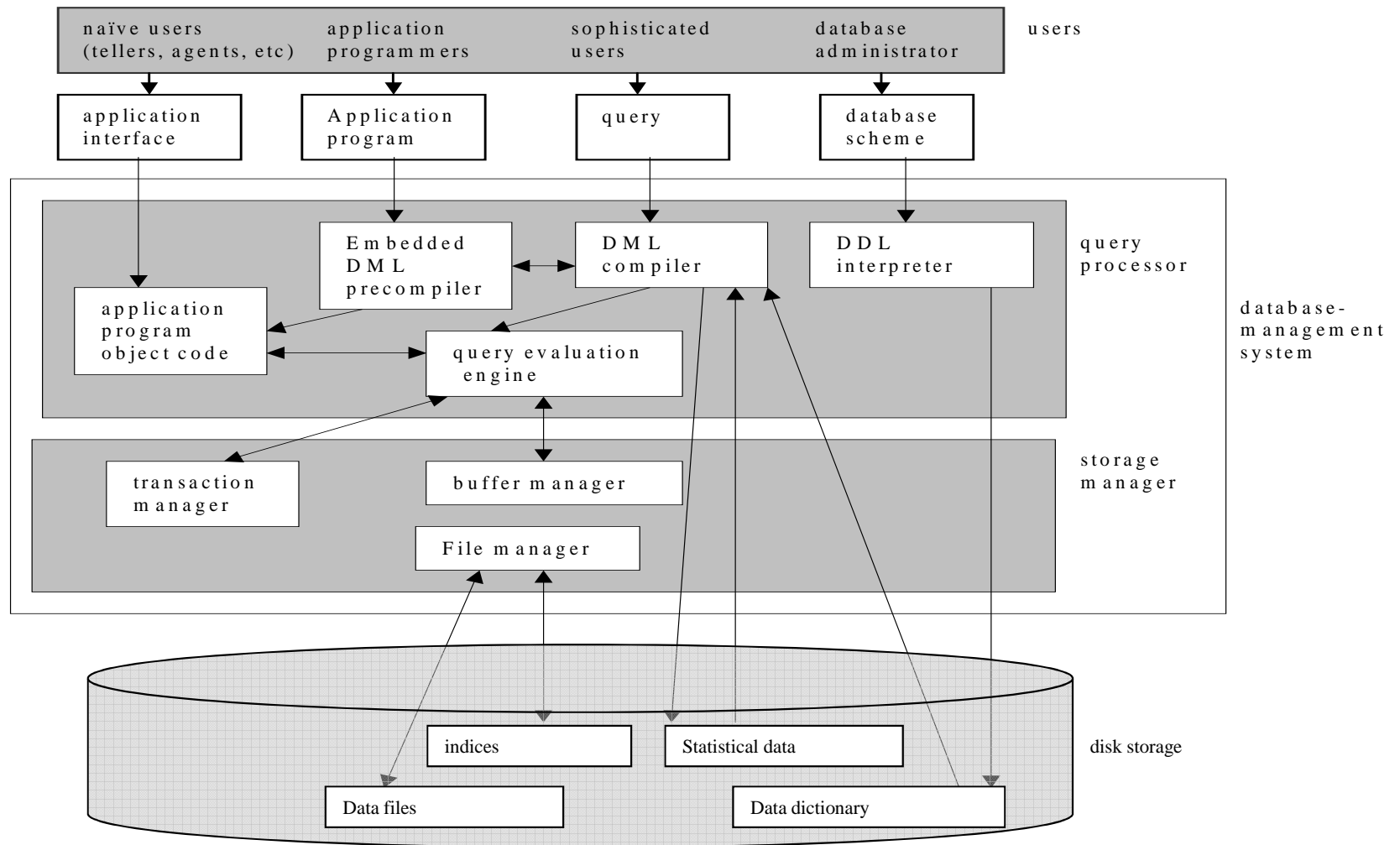
- Πολλές όψεις / προβολές (Views), μία μόνο εννοιολογική περιγραφή (conceptual (logical) schema) και μία μόνο φυσική περιγραφή (physical schema).
 - Οι ΟΨΕΙΣ περιγράφουν πως οι χρήστες βλέπουν τα δεδομένα
 - Το Conceptual schema ορίζει την λογική δομή των δεδομένων
 - Το Physical Schema περιγράφει την φυσική οργάνωση



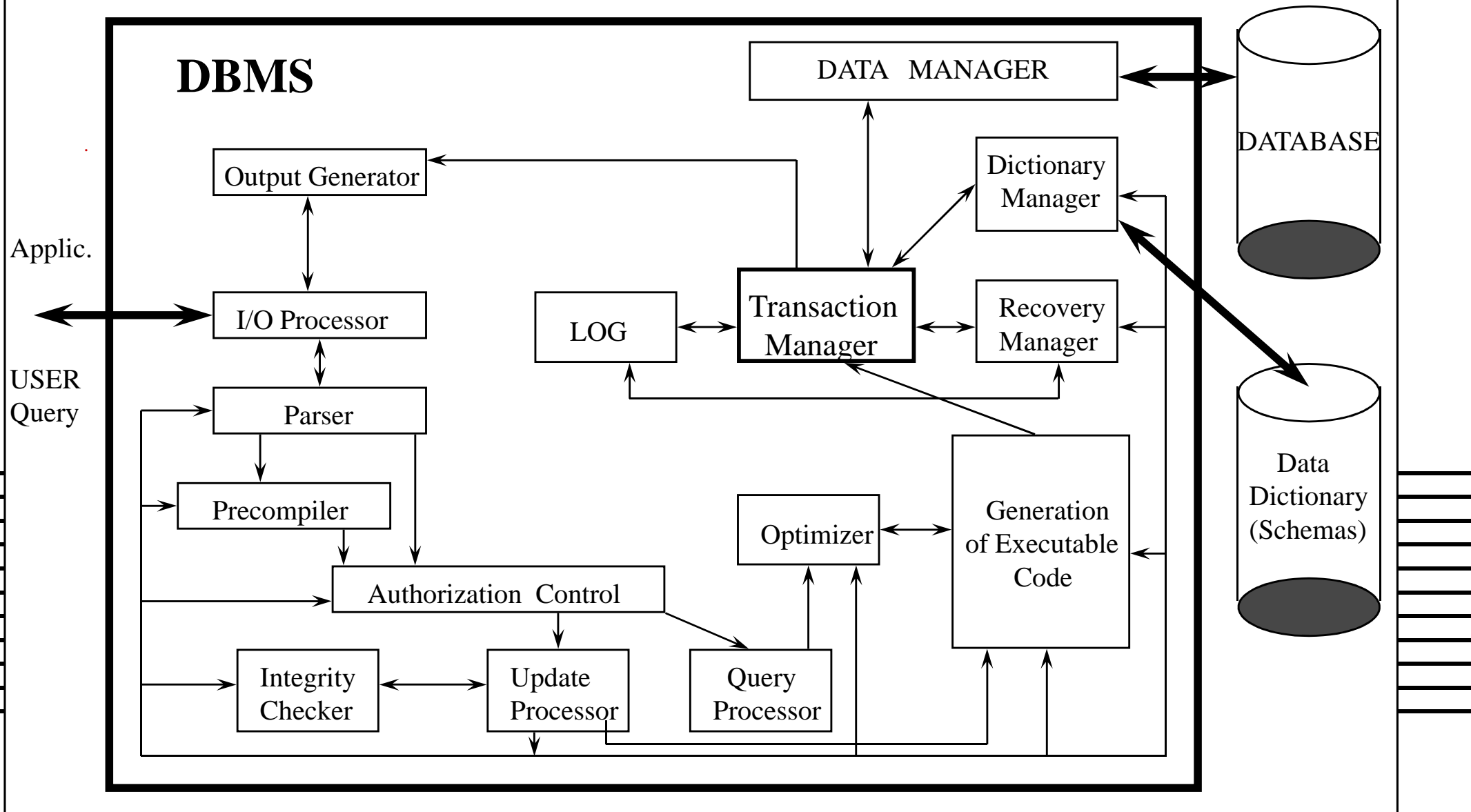
Παράδειγμα

- **Εννοιολογικό --- Conceptual schema (εννοιολογική διάταξη):**
*Students(sid: string, name: string, login: string,
age: integer, gpa:real)*
Courses(cid: string, cname:string, credits:integer)
Enrolled(sid:string, cid:string, grade:string)
- **Φυσικό --- Physical schema (φυσική διάταξη):**
 - Οι σχέσεις αποθηκεύονται ως ISAM αρχεία.
 - Ευρετήριο (Index) στην πρώτη στήλη του Students.
- **Εξωτερικό (Λογικό, Όψη, Προβολή) --- External Schema (View):**
Course_info(cid:string,enrollment:integer)

Γενική Αρχιτεκτονική ενός DBMS



Η Αρχιτεκτονική ενός DBMS



Έλεγχος Συνδρομικότητας - Concurrency Control

- Η ταυτόχρονη εκτέλεση δοσοληψιών / πράξεων σε ένα DBMS είναι ΑΠΑΡΑΙΤΗΤΗ για την ΚΑΛΗ ΑΠΟΔΟΣΗ.
 - Επειδή οι προσβάσεις στον Δίσκο είναι πολύ συχνές και σχετικά αργές, είναι πολύ σημαντικό να αφεθεί η CPU να εργάζεται σε πολλά προγράμματα χρηστών ΤΑΥΤΟΧΡΟΝΑ.
- Η τυχαία εναλλαγή μεταξύ προγραμμάτων των χρηστών μπορεί να οδηγήσει σε καταστροφή της ακεραιότητας της Βάσης
- Τα DBMS εξασφαλίζει ότι τέτοια φαινόμενα ΔΕΝ παρουσιάζονται – οι χρήστες παραμένουν με την εντύπωση ότι έχουν μοναδική χρήση του DBMS.

Δοσοληψία – Η εκτέλεση ενός Προγράμματος

- Η έννοια κλειδί είναι η Δοσοληψία (transaction), η οποία είναι μια ατομική σειρά πράξεων στη Βάση (reads/writes).
- Κάθε Δοσοληψία εκτελείται εντελώς και αφήνει την Βάση σε σωστή / συνεπή κατάσταση (consistent state) εφόσον ήταν σε συνεπή κατάσταση όταν άρχισε η Δοσοληψία.
- Οι χρήστες ορίζουν απλούς κανόνες ακεραιότητας (integrity constraints) στα δεδομένα, και το DBMS αναλαμβάνει να τους τηρεί
 - Πέραν αυτών το DBMS δεν καταλαβαίνει το νόημα των πράξεων. (π.χ., δεν καταλαβαίνει πως και γιατί υπολογίζεται ο τόκος σε έναν τραπεζικό λογαριασμό)

Transactions ACID

= μονάδες εργασίας που πρέπει να είναι:

1. *Atomic* = είτε εκτελούνται πλήρως ή καθόλου.
2. *Consistent* – Συνεπείς = οι σχέσεις μεταξύ των τιμών διατηρούνται.
3. *Isolated* – Απομονωμένες = φαίνονται να εκτελούνται σαν να μην υπάρχουν άλλες εκτελέσεις προγραμμάτων στο ΣΔΒΔ.
 - Συχνά αποκαλείται ως Σειριακή Συμπεριφορά
4. *Durable* – Ανθεκτικές = τα αποτελέσματα είναι μόνιμα, ακόμη και αν πέσει το σύστημα.

Λίγα Στοιχεία για τη Συνδρομικότητα

- Το DBMS εξασφαλίζει ότι η εκτέλεση των $\{T_1, \dots, T_n\}$ είναι ισοδύναμη (έχει το ίδιο αποτέλεσμα) με την ΣΕΙΡΙΑΚΗ (*serial*) εκτέλεση των $T_1' \dots T_n'$.
 - Πριν κάνει READ / WRITE σε μια εγγραφή (αντικείμενο της Βάσης) η Δοσοληψία ζητά το ΚΛΕΙΔΩΜΑ αυτού (LOCK), και περιμένει ώσπου να το πάρει. Όλα τα κλειδώματα ελευθερώνονται στο τέλος της δοσοληψίας
- (Strict 2PL locking protocol.)*
- **Ιδέα:** Αν κάποια πράξη του T_i (π.χ. WRITE X) επηρεάζει το T_j (π.χ., αυτό κάνει READ X), ένα από αυτά, ας πούμε το T_i , θα πάρει το κλειδί στο X πρώτο και το T_j αναγκάζεται να περιμένει μέχρι να τελειώσει το T_i , δηλαδή, μπαίνει μια σειρά στις Δοσοληψίες.
 - Τι συμβαίνει αν το T_j έχει ήδη ένα κλειδί στο Y και το T_i ζητά αργότερα ένα κλειδί στο Y? (**Deadlock!**)
Το T_i ή το T_j αναγκάζεται να εγκαταλείψει - **aborted** και να ξαναξεκινήσει!

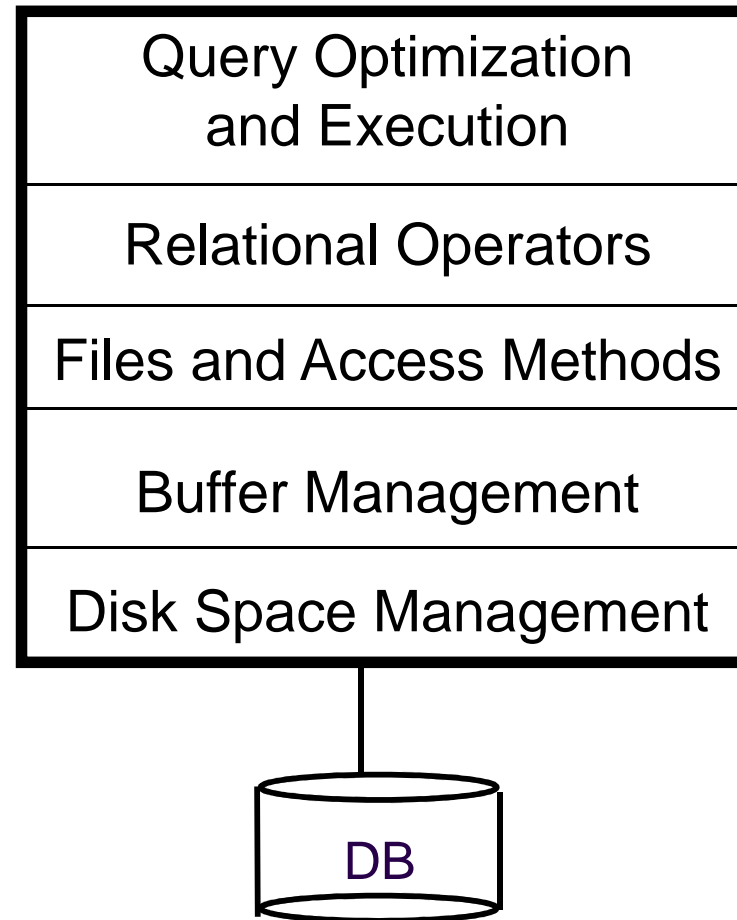
Ατομικότητα (Atomicity)

- Το DBMS διασφαλίζει την ατομικότητα (η ιδιότητα «ή όλα ή τίποτα») ακόμη και αν το σύστημα καταπέσει (*crash*) στην διάρκεια μιας πράξης.
 - Ιδέα: Κράτα ένα Ιστορικό (*log*) όλων των πράξεων του DBMS
 - Πριν γίνει μια αλλαγή στην Βάση, το ανάλογο ιστορικό φυλάσσεται σε ασφαλές σημείο. (WAL protocol)
- Οι μηχανισμοί του Λειτουργικού Συστήματος ΔΕΝ είναι αρκετοί .)
- Μετά το *crash*, τα μερικά αποτελέσματα μιας δοσοληψίας απαλείφονται (undone) με χρήση του *log*.

Απλούστευση της Αρχιτεκτονικής ενός DBMS

Αυτές τα στρώματα
Λαμβάνουν υπόψη
Concurrency
Control και Recovery

- Το τυπικό DBMS έχει μια σειρά από επίπεδα / στρώματα
- Αυτή είναι μια από τις πιθανές αρχιτεκτονικές – κάθε σύστημα μπορεί να έχει κάποια άλλη παραλλαγή αυτής.



ΟΙ ΧΡΗΣΤΕΣ ΤΩΝ DBMS

- **Διαχειριστές (Database Administrators)**
 - Υπεύθυνοι για την διαχείριση της Βάσης (βοηθούν χρήστες να ορίζουν όψεις, επιλέγουν εναλλακτικές δομές μνήμης και μεθόδους πρόσβασης, επίσης είναι υπεύθυνοι για την ασφάλεια και την ακεραιότητα της Βάσης, για την απόδοση του Συστήματος, κλπ.)
- **Σχεδιαστές (Database Designers)**
 - Υπεύθυνοι για τον Σχεδιασμό και Ανάπτυξη της Βάσης
- **Αναλυτές Συστημάτων και Προγραμματιστές (Application Programmers / Systems Analysts)**
 - Αναπτύσσουν τις Εφαρμογές
- **Τελικοί Χρήστες (End-Users)**
 - Κάνουν χρήσεις των εφαρμογών, κάνουν ερωταποκρίσεις, κλπ.

Περιστασιακοί (CASUAL USERS), Παραμετρικοί (PARAMETRIC) (π.χ., ταμίας τραπέζης – έτοιμα από την εφαρμογή) και Εξειδικευμένοι (SOPHISTICATED)

ΑΛΛΟΙ ΣΗΜΑΝΤΙΚΟΙ ΧΡΗΣΤΕΣ

- ***Database Designers and Implementers***
 - Αυτοί που αναπτύσσουν τα ΙΔΙΑ τα DBMS
- ***Tool Developers***
 - Αυτοί που αναπτύσσουν εργαλεία για καλύτερη χρήση των DBMS (design tools, performance tools, ειδικές προσβάσεις / διεπαφές.)
- ***Operators and Maintenance Personnel***
 - Βοηθούν τον Διαχειριστή, κάνουν συντήρηση, κλπ.

Γλώσσες σε ένα DBMS (1)

■ Γλώσσα Ορισμού των Δεδομένων --- Data Definition Language (DDL)

- Χρησιμοποιείται για να ορισθεί το *conceptual* schema της Βάσης – Αυτό αποθηκεύεται στον Κατάλογο (CATALOG). Συχνά, η DDL χρησιμοποιείται για τον ορισμό όψεων και φυσικών δομών. Σε ορισμένα DBMS γίνεται διαχωρισμός μεταξύ: "

SDL - Storage Definition Language (φυσικές δομές)

VDL - View Definition Language (όψεις)

■ Γλώσσα Επεξεργασίας Δεδομένων --- Data Manipulation Language (DML)

- Για την πρόσβαση (retrieval) και αλλαγές των δεδομένων (insert, delete, update)
- Δυο βασικές οικογένειες: ***Procedural DML, Declarative DML*** (*Διαδικαστικές και Δηλωτικές Γλώσσες*)

LANGUAGES ASSOCIATED WITH A DBMS (2)

- **Query Language**

- Ένα υποσύνολο μόνο για ερωταποκρίσεις - **RETRIEVAL**

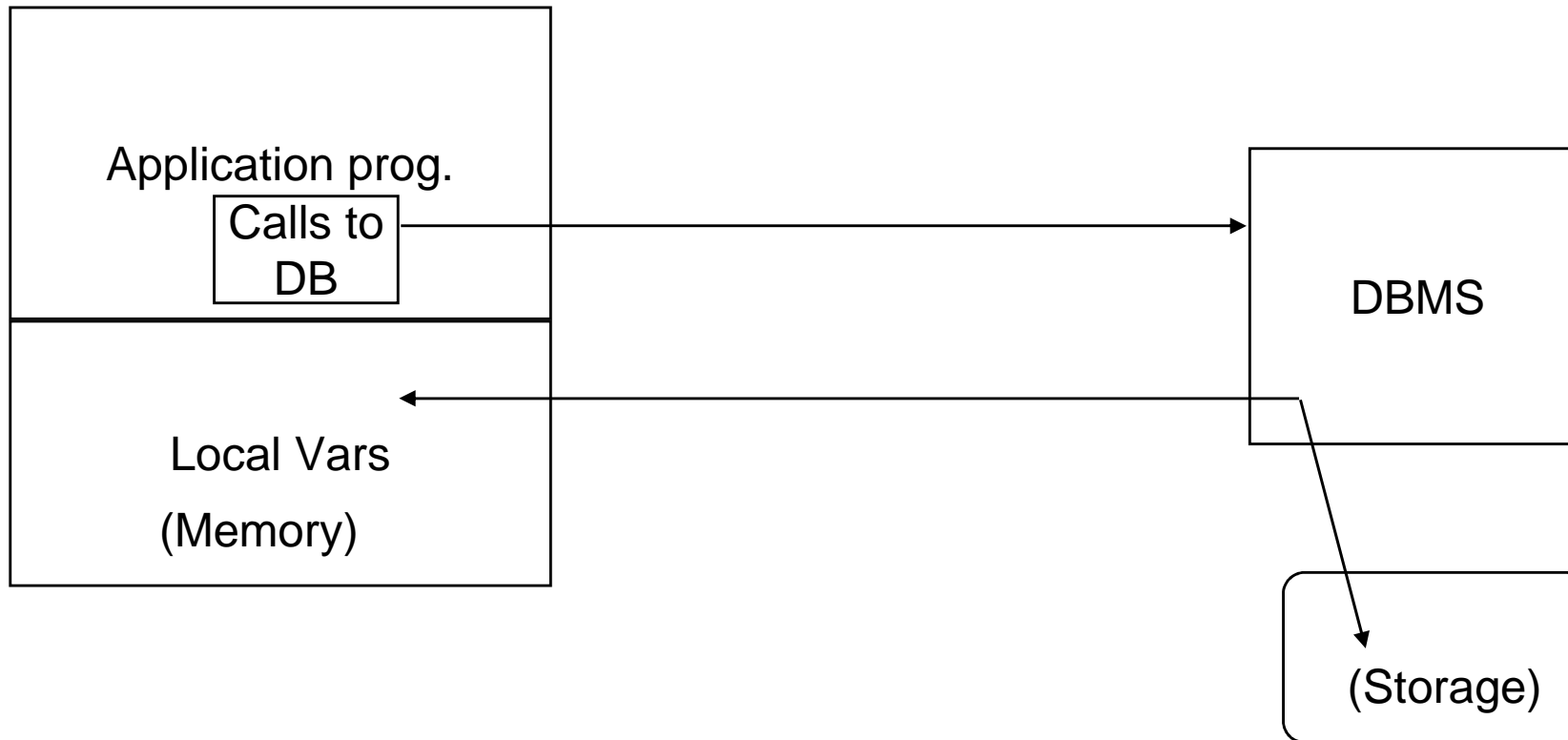
- **Data Sub-language**

- Η DML και η DDL μαζί.

- **Host Language**

- Μια ΓΛΩΣΣΑ ΠΡΟΓΡΑΜΜΑΤΙΣΜΟΥ (COBOL, C, etc..) όπου εμφυτεύονται εντολές της DML

Host Languages



- C, C++, Fortran, Lisp, COBOL, Java, ...
- Host language είναι μια Γενική Γλώσσα (Turing complete)

DBMS: Τρόποι Πρόσβασης

- **Αυτόνομη Query language Interface**
- **API (Application Programming Interface) για εμφυτευμένες Γλώσσες**
 - Pre-Compiler
 - Procedure Calls (subroutines)
- **Προσβάσεις μη-τεχνικών Χρηστών**
 - Menu-based, graphics-based, forms-based, φυσική γλώσσα, κλπ.
- **Παραμετρικές Προσβάσεις (function keys)**
- **Προσβάσεις για Αναφορές RGL)**
- **Πρόσβαση του Διαχειριστή - DBA**
 - Δημιουργία Λογαριασμών, Δικαιοδοσίες, κλπ
 - Ρύθμιση παραγόντων Συστήματος
 - Αλλαγές Περιγραφών, Δομών, κλπ.

Βοηθητικά Προγράμματα του DBMS (UTILITIES)

■ Λειτουργικότητα

- *Loading* δεδομένα από αρχεία στην Βάση
- *Backing up* τα δεδομένα σε Ταινίες
- *Reorganizing* Δομές της Φυσικής Βάσης
- *Report Generation Utilities*
- *Performance Monitoring Utilities*
- Άλλες Λειτουργίες (*sorting, user monitoring, data compression, κλπ.*)

■ Utilities του Καταλόγου Συστήματος

- Αποθήκευση των περιγραφών (schema descriptions), σχεδιαστικών αποφάσεων, προφίλ των χρηστών, τυποποιήσεις, περιγραφές εφαρμογών, κλπ.
- Ο ενεργός Κατάλογος είναι προσβάσιμος στους χρήστες και στο ίδιο το DBMS
- Το παθητικό λεξικό (data dictionary) είναι προσβάσιμο μόνο στους χρήστες

Βήματα για την Χρήση ενός DBMS

- **Προμήθεια του DBMS**
- **Εκπαίδευση Χρηστών**
- **Ορισμός των περιγραφών (Schemas) της Βάσης**
- **Εισαγωγή Δεδομένων στη Βάση**
- **Υλοποίηση των Προγραμμάτων της Εφαρμογής**
- **Συνεχείς ΑΛΛΑΓΕΣ στη Βάση και την Εφαρμογή**

Γιατί ΝΑ ΜΗΝ χρησιμοποιήσω DBMS ;;;

■ Κόστος

- Υψηλότετη αρχική επένδυση
- Τεράστιο Πρόσθετο Κόστος (Overhead) για την ασφάλεια, ανάκαμψη ακεραιότητα, και συνδρομικότητα (κόστος στην απόδοση του Συστήματος).

■ Πότε ΔΕΝ Χρειάζεται για την εφαρμογή

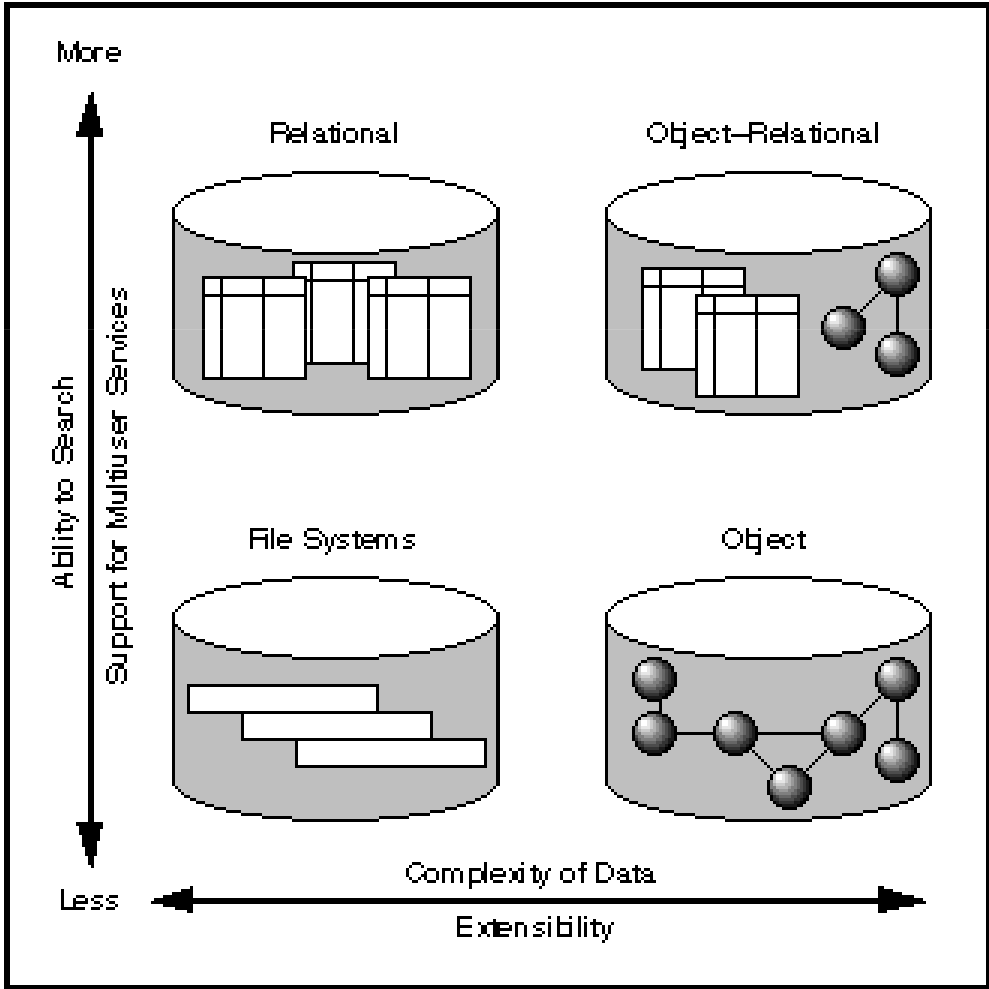
- Η Βάση Δεδομένων μικρή και η Εφαρμογές είναι πολύ απλές και σταθερές (σε τέτοια περίπτωση, οι επιδόσεις – λαμβάνοντας υπόψη και το overhead του DBMS – πιθανόν να είναι χειρότερες μιας απλής υλοποίησης)
- - Δεν απαιτείται πρόσβαση από πολλούς χρήστες

Ιστορική Αναδρομή (4)

■ 1990-2000s (Πέμπτη Γενιά – Μετά-Σχεσιακή Εποχή)

- Η εποχή των ΣΥΝΘΕΤΩΝ ΟΝΤΟΤΗΤΩΝ σε Βάσεις Δεδομένων (*engineering objects, πολυμέσα - multimedia, software objects*)
 - **Object-Relational Database Systems**
- Ενεργές (Active) Βάσεις Δεδομένων, Intelligent Systems, Πολυμέσα
- Εξυπηρετητές / πελάτες
- Αποθήκες Δεδομένων (Data Warehouse)
- Εξόρυξη Δεδομένων / Πληροφοριών (Data Mining)
- Κατανεμημένες (Multidatabases), Παραλληλοποίηση
 - » Συστήματα σε Προσωπικούς Υπολογιστές
 - » Συστήματα στο Internet (Web-based), Java, XML, ..., **Cloud**

Σχεσιακές και Αντικειμενοστραφείς Βάσεις



Πρώτη Γενιά Πελάτη-Εξυπηρετητή

ΠΕΛΑΤΗΣ

Υπηρεσίες Παρουσίασης
Λογική παρουσίασης
Λογική Εφαρμογής
Επιχειρησιακοί Κανόνες
Λογική Δεδομένων

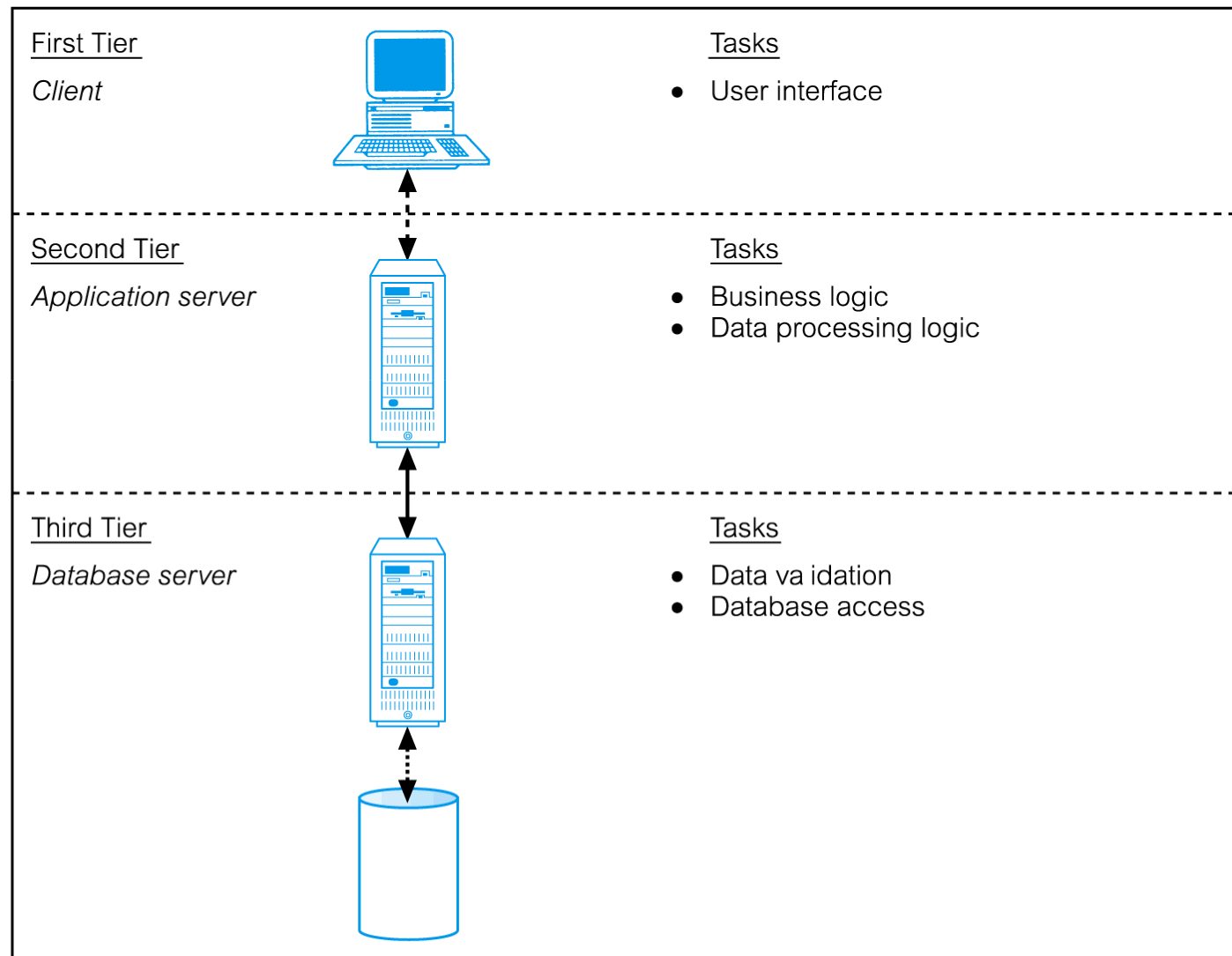
ΕΞΥΠΗΡΕΤΗΤΗΣ

Υπηρεσίες
Βάσεων
Δεδομένων

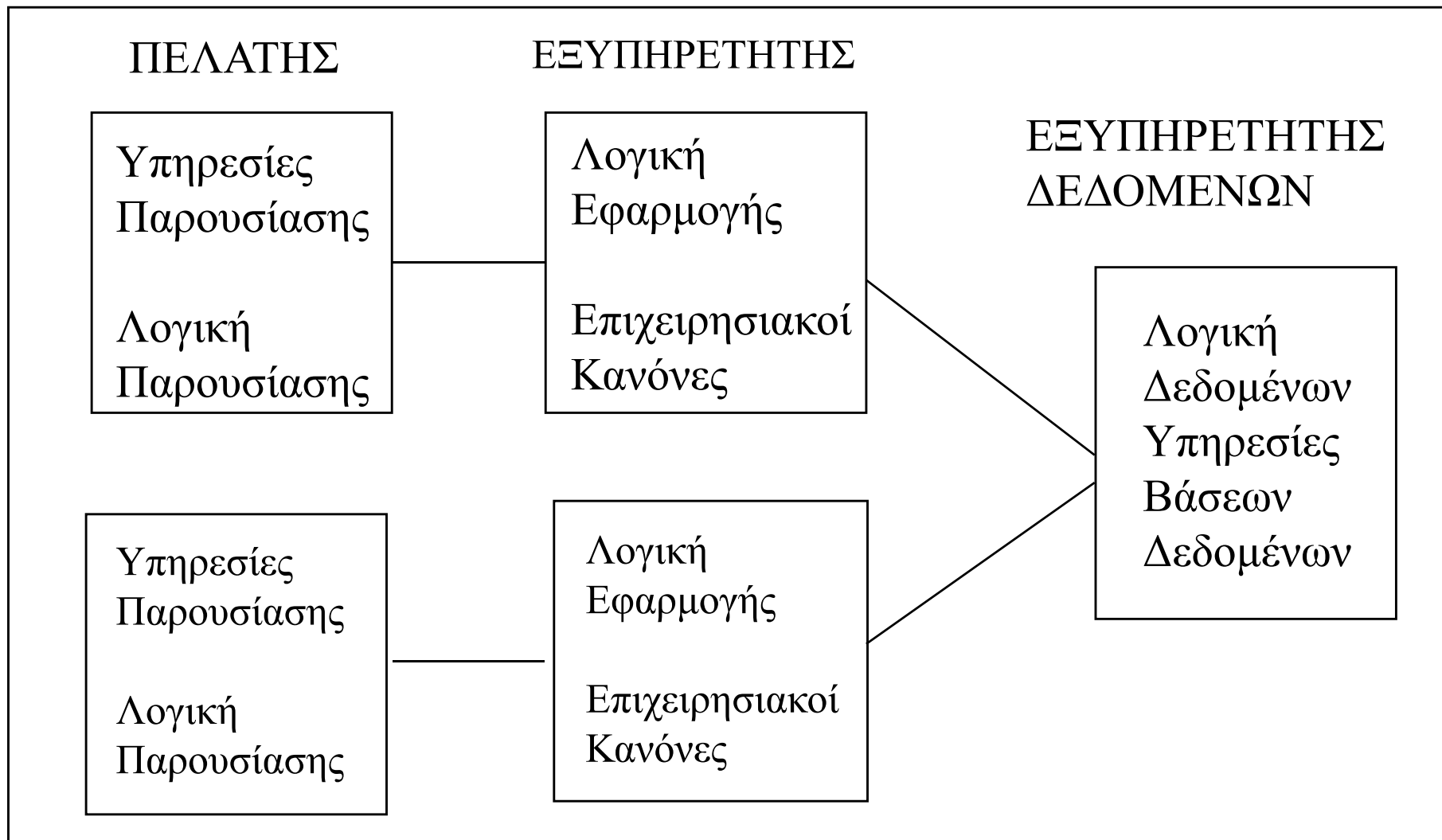
Αρχιτεκτονικά στοιχεία

- **Υπηρεσίες Παρουσίασης**
 - Παίρνει εισόδους και παρουσιάζει αποτελέσματα
- **Λογική Παρουσίασης**
 - Ελέγχει την σχέση μεταξύ χρήστη - εφαρμογής
- **Λογική Εφαρμογής**
 - Υπολογισμοί, αποφάσεις, δράσεις για την εφαρμογή
- **Επιχειρησιακοί Κανόνες**
 - Ολόκληρης της Επιχείρησης
- **Λογική Δεδομένων**
 - Δημιουργία μιας ερώτησης (π.χ. στην SQL)
- **Υπηρεσίες Βάσεων Δεδομένων**
 - Εξυπηρετεί τις αναφορές σε δεδομένα

Αρχιτεκτονική 3 Επιπέδων – 3-Tier



Δεύτερη Γενιά Client-Server



Υλικό και Λειτουργικό Σύστημα

■ Πελάτης

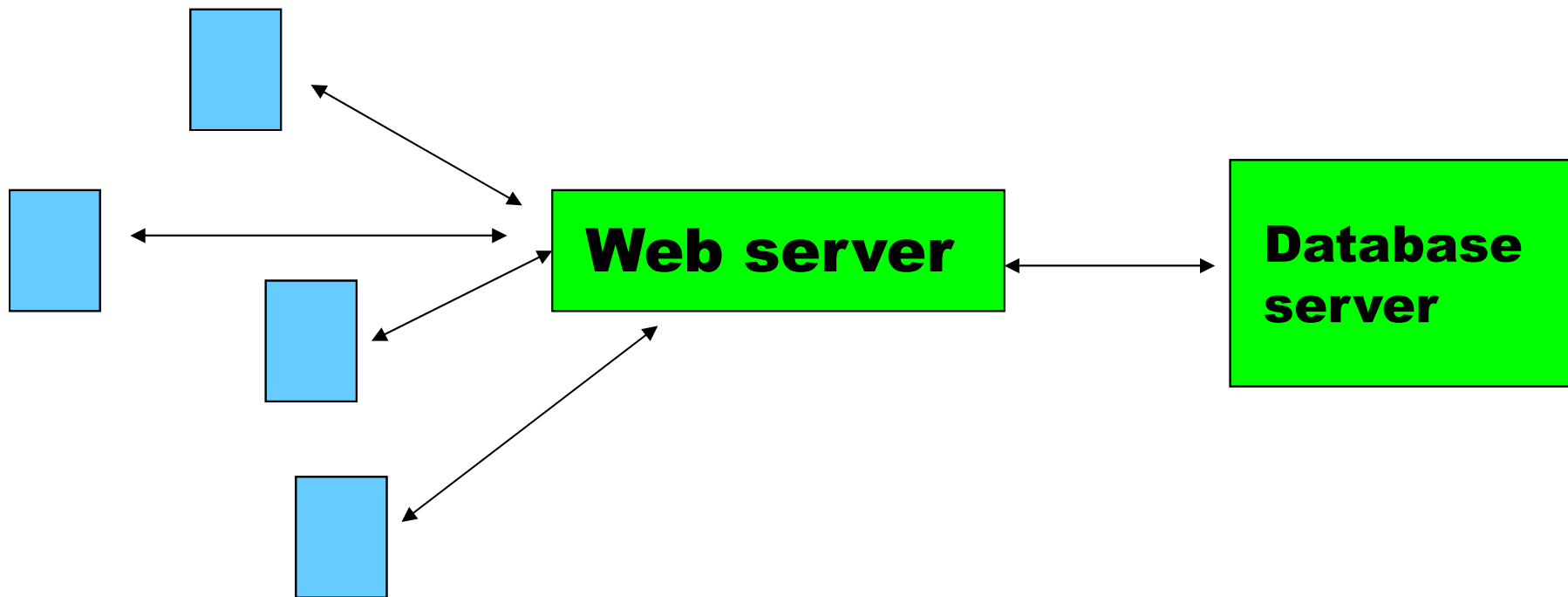
- Ένας επεξεργαστής
- PC (π.χ., PowerPC, Power Mac)
- Windows XP, ... NT, ... 2000, Apple Mac, OS/2, Unix

■ Εξυπηρετητής

- 2 έως 100 χρήστες
- μέχρι μεγάλες μηχανές (SMP, clusters, MPP)
- Windows NT, 2000, Novell Netware, Unix

Σημερινή Τυπική Διάταξη - Αρχιτεκτονική

Clients



Τεχνολογία Διοίκησης Δεδομένων

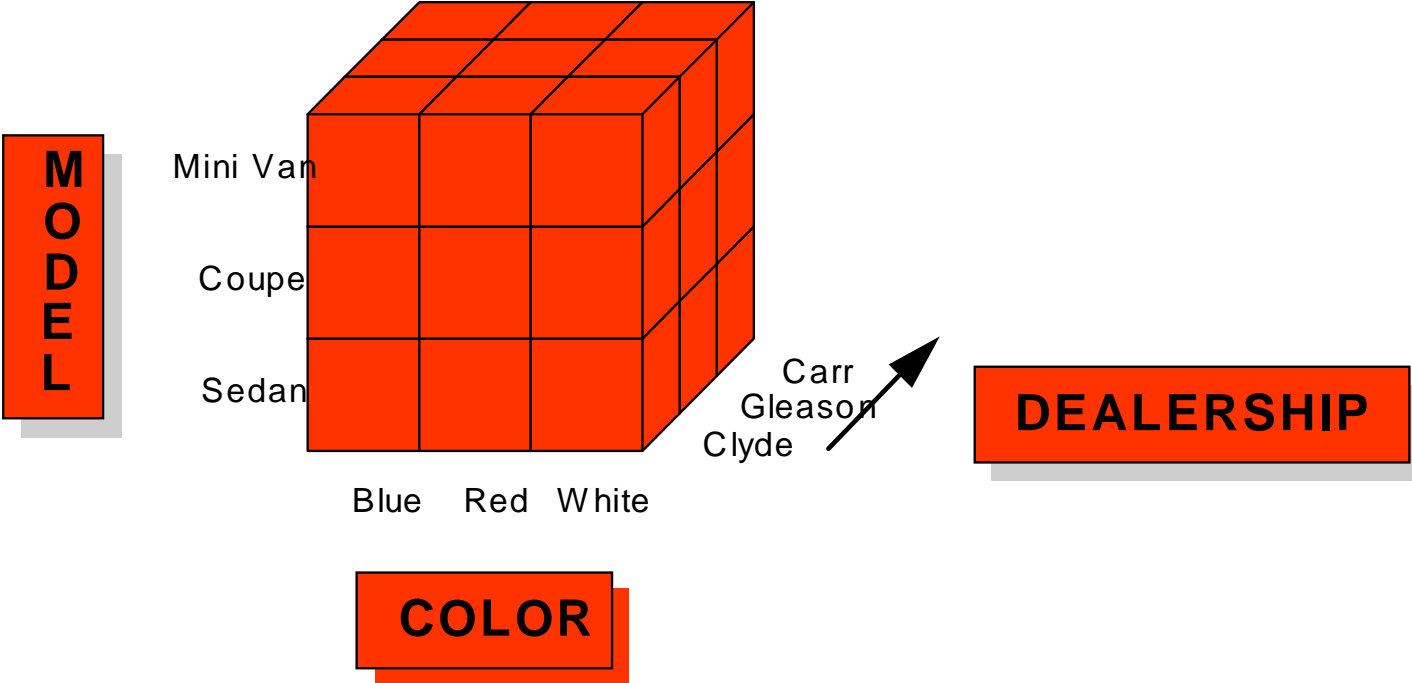
- **OLAP (Online Analytic Processing) – Αναλυτική Επεξεργασία**
 - Ανάλυση σε πολυδιάστατο χώρο
 - » Πωλήσεις ANA (Προϊόν, Πελάτη, Χρόνο)
 - » Κύβος
 - drill down, rollup
 - Essbase, Commander, Oracle Express, SAS, Excel, SQL Server,...
- **Data Warehouses**
 - Νέα μορφή για Decision Support Systems (Συστήματα Στήριξης Αποφάσεων)
 - Redbrick, Oracle DW, Informix, Sybase, Micro Strategy
- **Data Marts (Μικρότερα Data Warehouses)**
- **Data Mining (Εξόρυξη Δεδομένων – Πληροφοριών)**

OLAP - Εφαρμογές

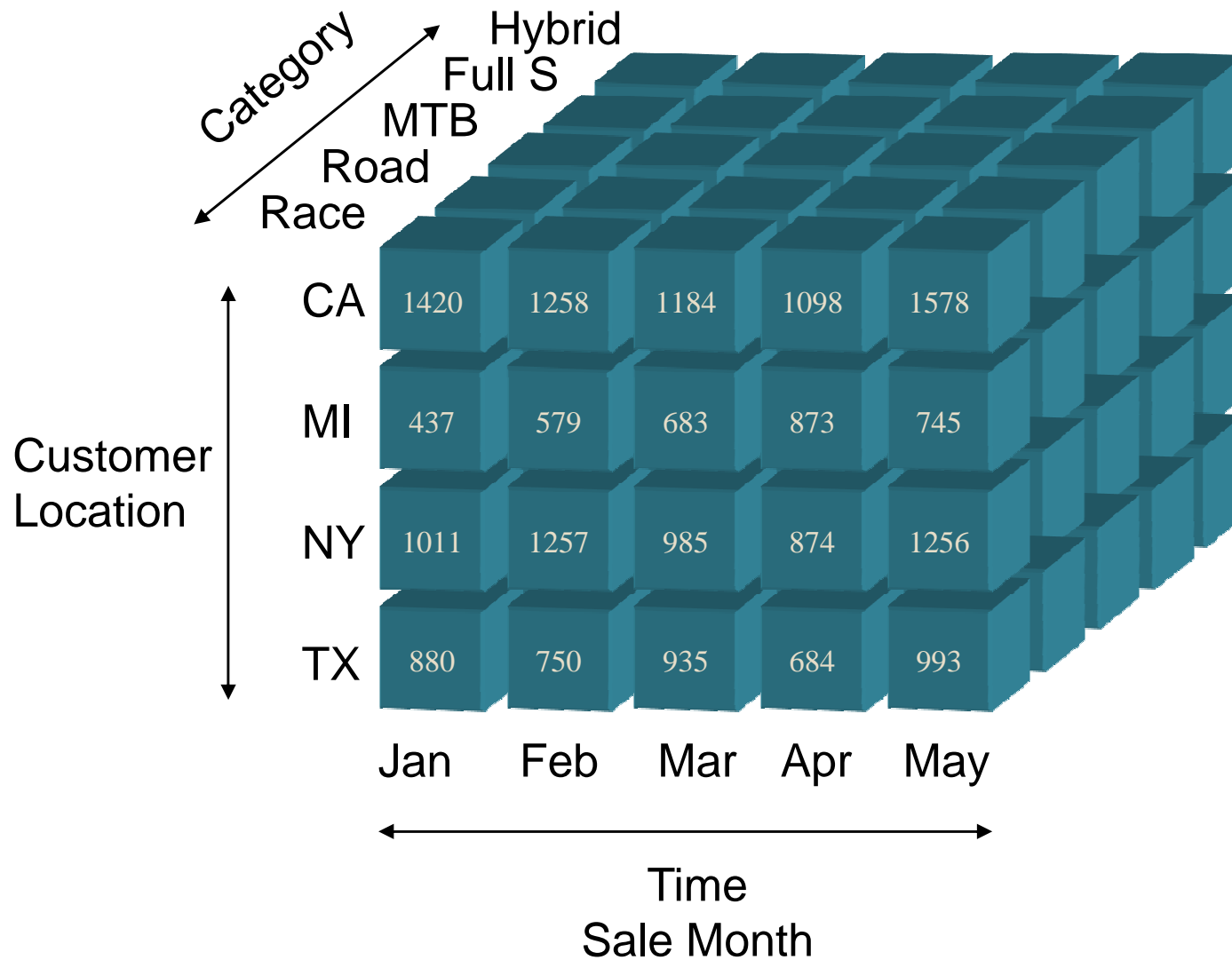
- **Κλασσικές Εφαρμογές**
 - Οικονομική Ανάλυση, Μοντελοποίηση και Αναφορές
 - Προϋπολογισμός,
 - Διασφάλιση Ποιότητας και Έλεγχος Ποιότητας,
 - Κερδοφορία,
 - Ανάλυση Έρευνας
 - Παρακολούθηση προώθησης προϊόντων
- **OLAP και OLTP δεν μπορούν να δουλέψουν ταυτόχρονα στα ίδια λειτουργικά δεδομένα!**
 - εντελώς διαφορετικές και συγκρουόμενες απαιτήσεις

OLAP: Η έννοια του KYBOY

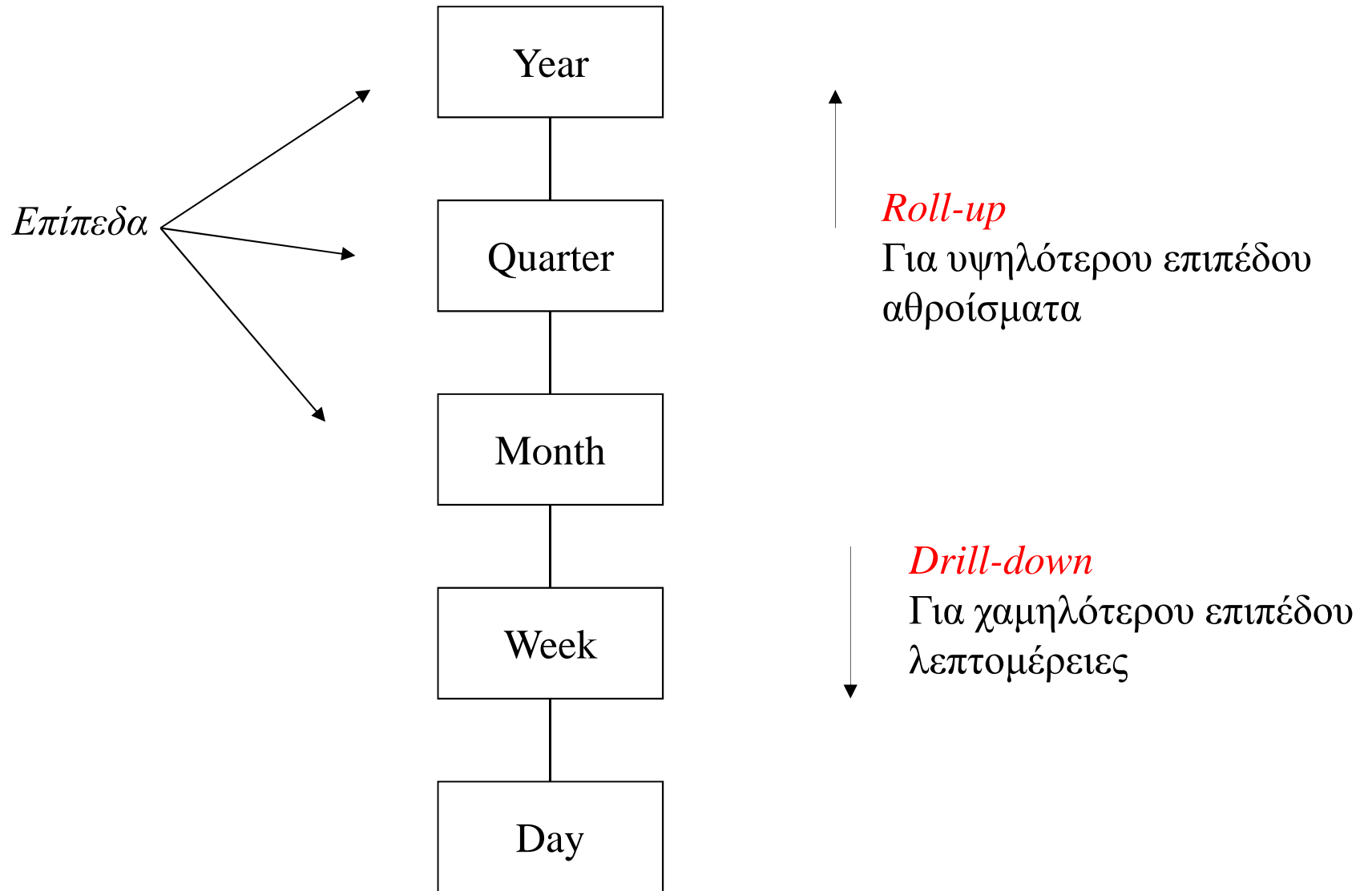
Sales Volumes



Πολυδιάστατος Κύβος - Παράδειγμα



Βασικές Πράξεις / Περιήγηση

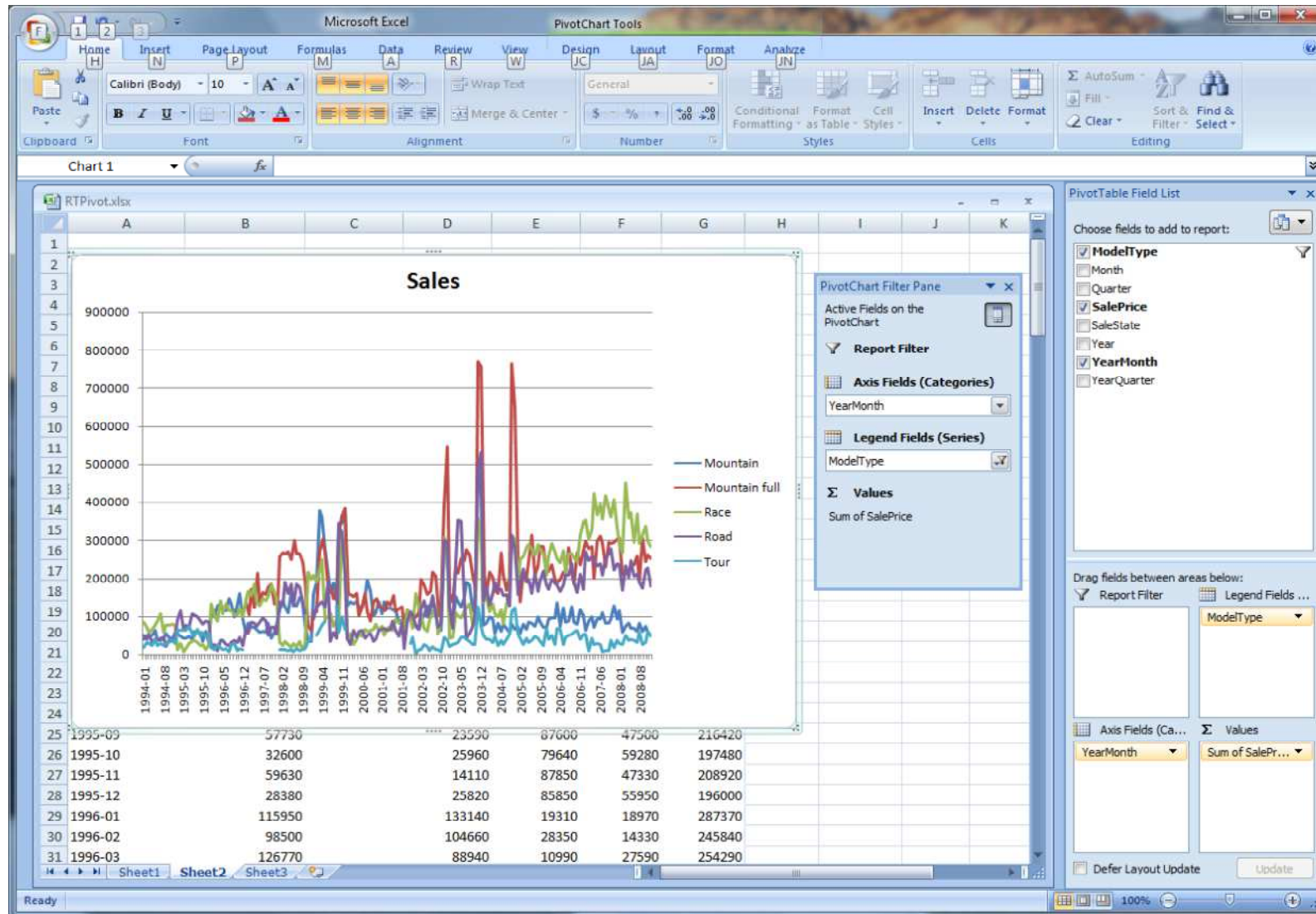


Microsoft Pivot Table

The screenshot displays Microsoft Excel with a PivotTable and the PivotTable Field List task pane. The PivotTable summarizes sales data by year and month, categorized by model type (Mountain, Road, Tour) and Grand Total. The PivotTable Field List shows the current configuration: Report Filter (ModelType), Row Labels (Year, YearMonth), and Values (Sum of SalePrice).

Row Labels	Mountain	Mountain full	Race	Road	Tour	Grand Total
1994	\$559,020		\$928,984	\$486,774	\$346,201	\$2,320,978
1995	\$567,940		\$294,390	\$1,074,490	\$709,230	\$2,646,050
1996	\$1,469,870		\$1,374,850	\$361,020	\$213,360	\$3,419,100
1997	\$783,820	\$1,826,320	\$1,851,160	\$894,040		\$5,355,340
1998	\$1,366,180	\$2,712,310	\$683,090	\$1,562,180	\$203,270	\$6,527,030
1999	\$3,138,210	\$2,792,330	\$1,998,490	\$1,816,790	\$634,260	\$10,380,080
2000	\$1,657,070	\$1,611,920	\$671,200	\$600,180		\$4,540,370
2001	\$1,221,870	\$1,399,290	\$936,430	\$876,740	\$72,890	\$4,507,220
2002	\$1,523,740	\$2,507,560	\$1,581,050	\$1,645,770	\$225,980	\$7,484,100
2003	\$2,241,600	\$3,675,790	\$1,585,560	\$2,593,660	\$526,680	\$10,623,290
2004	\$896,390	\$3,372,690	\$1,858,720	\$2,183,070	\$624,800	\$8,935,670
2005	\$934,540	\$2,812,930	\$3,113,150	\$2,285,410	\$486,990	\$9,633,020
2006	\$1,205,720	\$2,700,080	\$3,235,570	\$2,319,750	\$599,460	\$10,060,580
2007	\$1,098,330	\$3,370,890	\$4,431,690	\$2,965,450	\$313,340	\$12,179,700
2007-01	\$63,590	\$297,020	\$354,110	\$270,920	\$54,250	\$1,039,890
2007-02	\$88,660	\$278,430	\$303,620	\$247,930	\$9,510	\$928,150
2007-03	\$98,780	\$282,460	\$326,320	\$259,260	\$42,830	\$1,009,650
2007-04	\$71,210	\$199,680	\$424,210	\$247,070	\$26,860	\$969,030
2007-05	\$92,080	\$297,740	\$367,540	\$234,630	\$30,140	\$1,022,130
2007-06	\$93,320	\$310,440	\$393,820	\$238,340	\$25,430	\$1,061,350
2007-07	\$109,880	\$292,690	\$357,180	\$209,070	\$8,360	\$977,180
2007-08	\$83,090	\$228,970	\$418,350	\$252,440	\$11,320	\$994,170
2007-09	\$82,370	\$293,490	\$382,680	\$265,360	\$11,160	\$1,035,060
2007-10	\$93,430	\$292,910	\$363,300	\$278,830	\$38,260	\$1,066,730
2007-11	\$128,150	\$295,100	\$407,050	\$223,210	\$18,670	\$1,072,180
2007-12	\$93,770	\$301,960	\$333,510	\$238,390	\$36,550	\$1,004,180
2008	\$846,990	\$2,930,500	\$3,931,900	\$2,508,860	\$471,980	\$10,690,230
Grand Total	\$19,511,290	\$31,712,610	\$28,476,234	\$24,174,184	\$5,428,441	\$109,302,758

Microsoft Pivot Chart



Σχεδιασμός Νιφάδας Χιονιού

Merchandise
ItemID
Description
QuantityOnHand
ListPrice
Category

Sale
SaleID
SaleDate
EmployeeID
CustomerID
SalesTax

City
CityID
ZipCode
City
State

Customer
CustomerID
Phone
FirstName
LastName
Address
ZipCode
CityID

OLAPItems
SaleID
ItemID
Quantity
SalePrice
Amount

Σχεδιασμός Αστεριού

Dimension Tables

Products

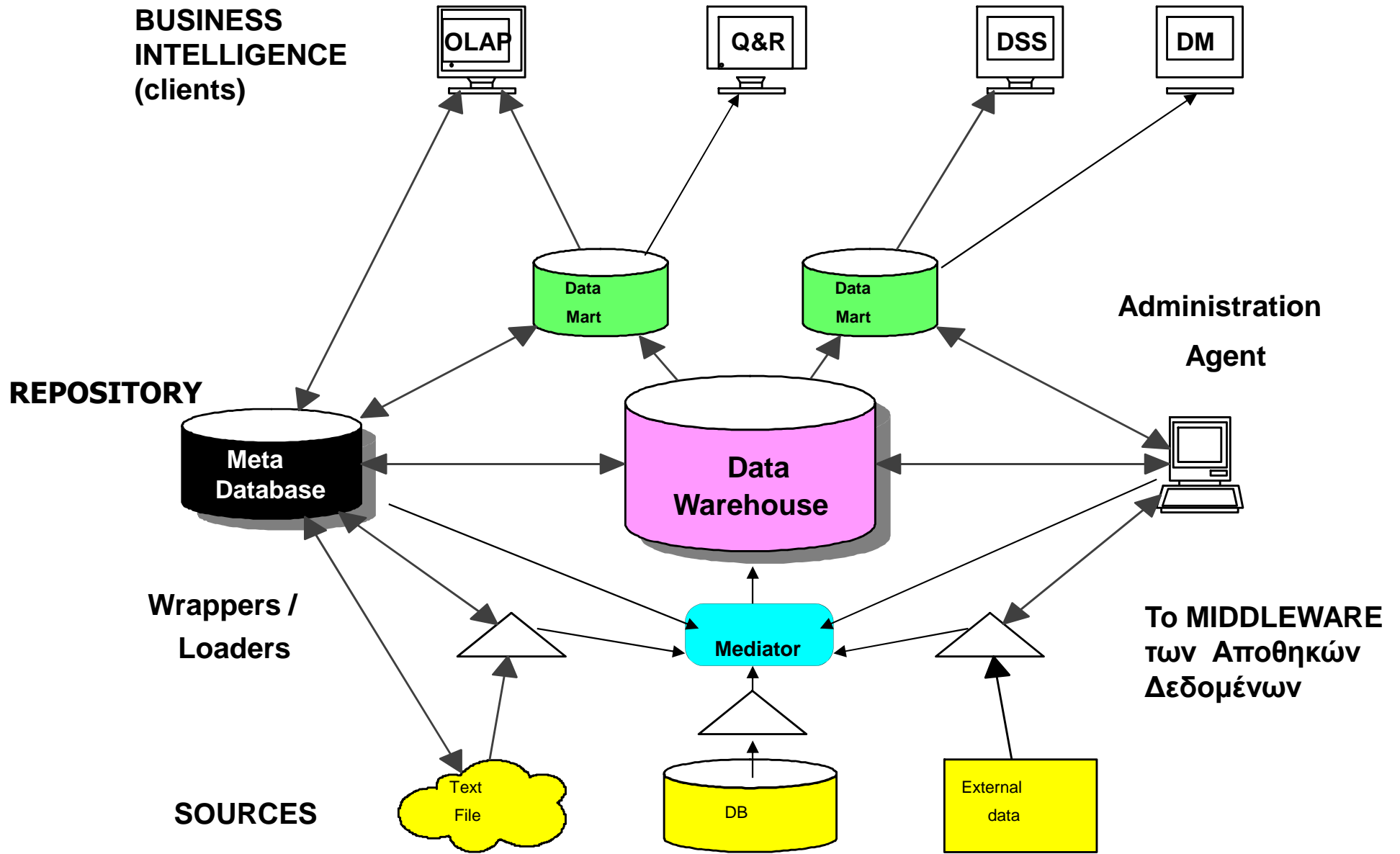
Sales Date

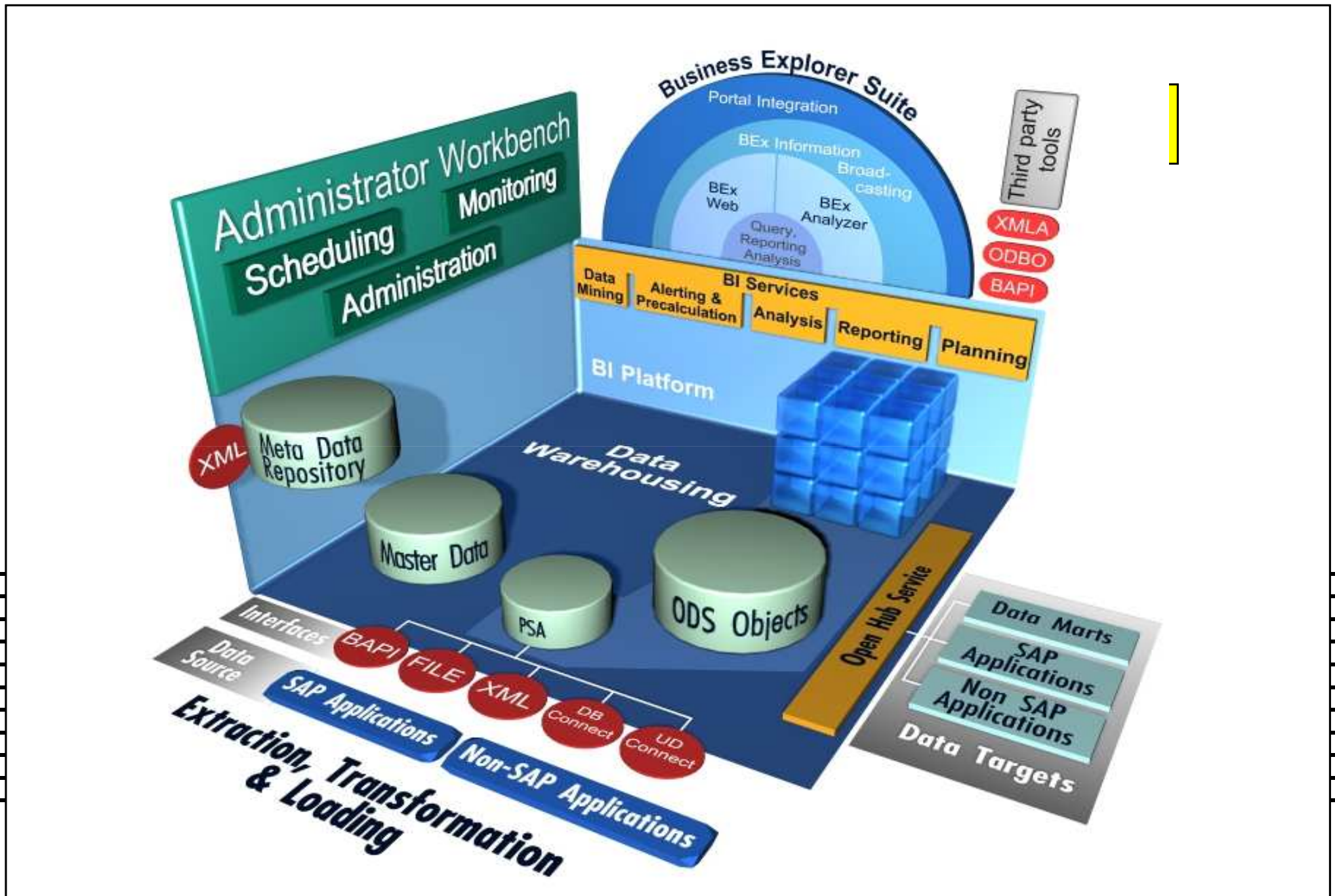
Fact Table

Sales
Quantity
Amount=SalePrice*Quantity

Customer
Location

Κλασική Αρχιτεκτονική DW





Ανάλυση Πωλήσεων Ανά Χώρα

BW Web A

File Edit View Favorites Tools Help

Address http://sapteka8:8000/sap/bw/BEx?SAP-LANGUAGE=E&PAGENO=7&CMD=PROCESS_VARIABLES&REQUEST_NO=1&CMD=PROCESS_VARIABLES&SUBCMD=VAR_SUBMIT&VARID=

Google Search

Calendar Month/Year JAN 2004..MAR 2004

SALES BY COUNTRY query

Conditions: Top 5 Active

Create

Legend:

- GREECE
- UNITED KINGDOM
- ITALY
- FRANCE
- GERMANY

SALES BY COUNTRY query

Country of the payer	Sales amount	Weight (KG)
GR GREECE	16.893.753,55 EUR	8.213.775 KG
GB UNITED KINGDOM	3.162.383,68 EUR	1.726.927 KG
IT ITALY	1.495.587,27 EUR	496.400 KG
FR FRANCE	1.426.217,46 EUR	661.455 KG
DE GERMANY	863.724,14 EUR	356.021 KG
Overall Result	28.469.312,63 EUR	13.640.028 KG

Navigation Block:

Bill-to party			
Country of the payer			
Division			
Material			
Sales Office			
Key Figures			

Local intranet

start BW κη... Inbox ... SAP L... Micros... Micros... BEx W... SAP E... BW W... 6:25 μμ

Καρτέλα Εργαζομένου

BW Web Application - Microsoft Internet Explorer provided by Vodafone Greece

Address: http://grxha112.gr.sedc.internal.vodafone.com:8000/sap/bw/BEx?SAP-LANGUAGE=E&PAGENO=2&CMD=PROCESS_VARIABLES&REQUEST_NO=1&CMD=PROCESS_VARIABLES&SUBCMD=VAR_SUBMI

Employee Information

Employee	Entry Date	Organizational Unit	Job	Scale ID	Band / Level	Gender	Age in Years	Length of Service	Gross Salary
Last Name00003717 First Name00003717	02.11.2004	BUSINESS SYSTEMS INT	50029221	ACC MGT	2	Female			6.300,00 EUR
				#		Female	105,00	1,00	

Personal Information

Address	City	Postal Code	Home Phone	Office Phone	Marital Status
ΔΙΕΥΘΥΝΣΗ ΚΗΦΙΣΙΑΣ 44	ΠΟΛΗ ΜΑΡΟΥΣΙ	11111	TEL00003717	6567	Marr.

Photo

Salary Comparison

Category	Value
Employee Gross Salary	6300 Euro
Average Salary of BUSINESS SYSTEMS INT	3308,63 Euro

Service Comparison

Category	Value
Employee Length of Service	1,00 Years
Average Service of BUSINESS SYSTEMS INT	4,58 Years

Education

Level	Institute	Country	Title	Certificate	Duration
AEI	ΠΑΝΤΕΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ	Greece	Political Science	Final certificate	4 Years

Foreign Languages

Title	Institute	Country	Certificate
ΑΓΓΛΙΚΑ - Proficiency	British Council	Greece	Final certificate

- > [Salary History](#)
- > [Overtime History](#)
- > [Stand By History](#)
- > [Shifts History](#)

Done Local intranet

Planning & Budgeting

Planning Edit Goto Utilities Tools System Help

Enter plan data

Close navigation Planning profile Global planning sequences Set Variables

Planning areas	Technical name
Actual & Budget data	SALES03
ΕΞΩΤΕΡΙΚΟ	PL01
Copy actual to budget	PP01
ΕΣΩΤΕΡΙΚΟ	PL02
Copy actual to budget	PP01
Sales Budget	SALES02
Εισαγωγή budget - ΕΞΩΤΕΡΙΚΟ	PL03
Budget	PP01
Εισαγωγή budget - ΕΣΩΤΕΡΙΚΟ	PL04
Budget	PP02
Υπολογισμός αξίας	PL05
Budget price X Actual Qty	PP01

Planning functions	Technical name
Εισαγωγή budget - ΕΞΩΤΕΡΙΚΟ	PL03
Manual planning	0-MP
Manual Planning	MP01

ΦΑΣΟΝ # Not assigned
Fiscal year/period 001.2005 January 2005

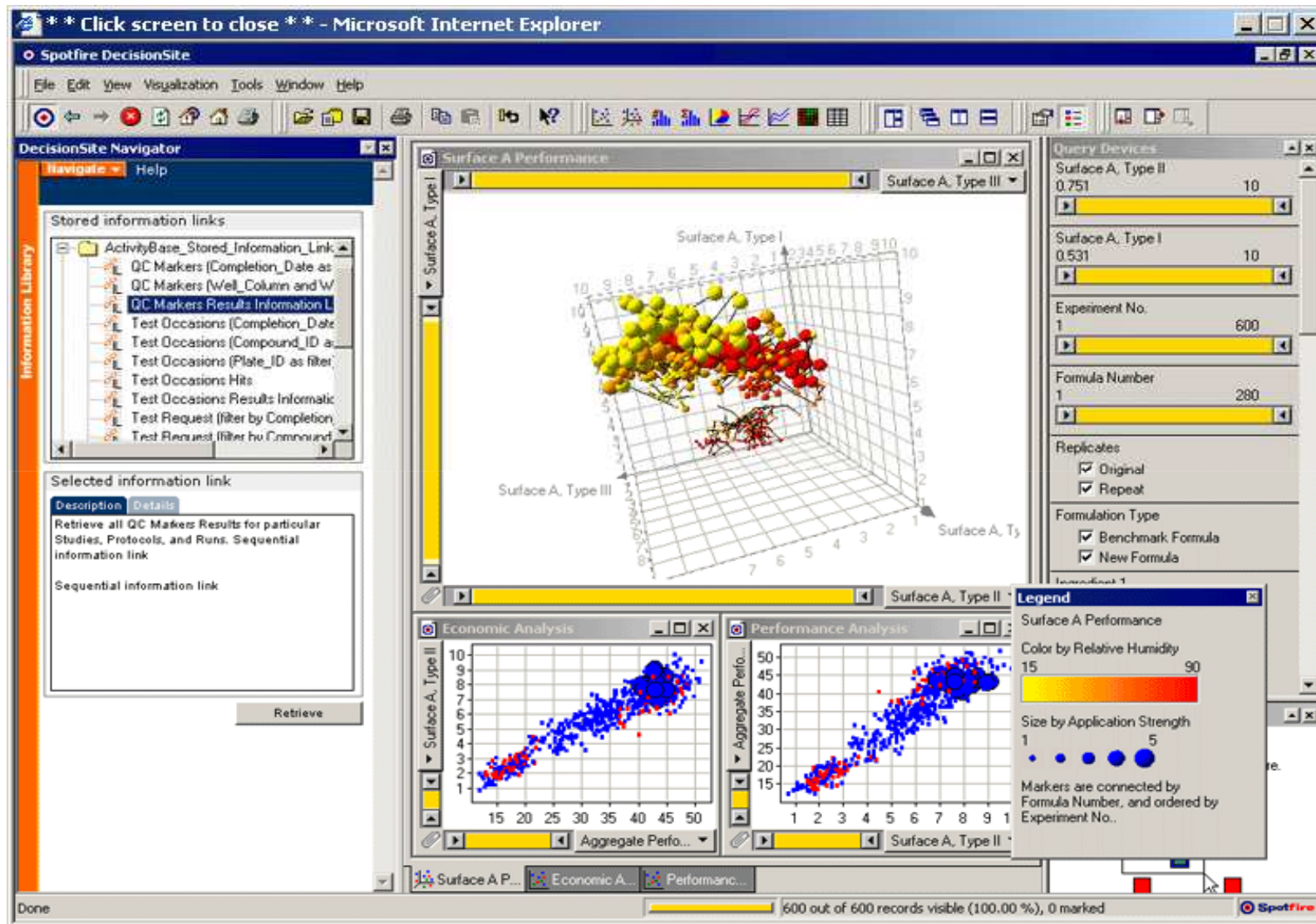
Material group		E.E. (Ton)	E.E. (Τιμή)	Currency	ΤΡΙΤΕΣ ΧΩΡΕΣ (Ton)	ΤΡΙΤΕΣ ΧΩΡΕΣ (Τιμή)	Currency
01010	SH...	72,000	420,00	EUR	0,000	0,00	USD
01020	SH...	20,000	590,00	EUR	62,000	645,00	USD
01030	SH...	142,000	540,00	EUR	794,000	710,00	USD
01040	SH...	65,000	700,00	EUR	389,000	970,00	USD
01050	SH...	122,000	800,00	EUR	457,000	1.120,00	USD
01054	SH...	132,000	1.000,00	EUR	231,000	1.300,00	USD
01060	SH...	8,000	900,00	EUR	0,000	0,00	USD
01064	SH...	21,000	1.550,00	EUR	0,000	0,00	USD
01070	SH...	3,000	460,00	EUR	0,000	0,00	USD
01080	SH...	51,000	450,00	EUR	0,000	0,00	USD
01090	SH...	0,000	0,00	EUR	0,000	0,00	USD
01100	SH...	56,000	600,00	EUR	0,000	0,00	USD
01110	SH...	3,000	800,00	EUR	25,000	900,00	USD
01120	SH...	3,000	800,00	EUR	198,000	900,00	USD
01130	SH...	0,000	0,00	EUR	34,000	1.120,00	USD
01200	SH...	17,000	750,00	EUR	0,000	0,00	USD
02010	CO...	42,000	325,00	EUR	0,000	0,00	USD
02020	CO...	173,000	520,00	EUR	225,000	500,00	USD
02030	CO...	49,000	615,00	EUR	17,000	670,00	USD
02050	CO...	241,000	1.500,00	EUR	38,000	1.885,00	USD
02054	CO...	303,000	1.450,00	EUR	0,000	0,00	USD

BWP (1) (900) sapteka5 INS

Data Mining

- *Αυτόματη Ανάλυση Δεδομένων*
- *Στατιστικά*
 - Correlation
 - Regression (multiple correlation)
 - Clustering
 - Classification
 - Nonlinear relationships
- *Επιπλέον Αυτοματισμοί*
 - Ανάλυση του καλαθιού αγοράς
- *Αριθμητικά Δεδομένα και μη-αριθμητικά δεδομένα*
 - Γλωσσική Ανάλυση

Data Mining Εργαλεία: Spotfire



<http://www.spotfire.com>

Market Basket Analysis

Τι αγοράζουν «μαζί» οι Πελάτες?



Data Mining: Market Basket Analysis

- **Στόχος: Μέτρηση της σχέσης μεταξύ δύο ειδών**
 - Τι είδη αγοράζουν οι πελάτες μαζί?
 - Ποιες σελίδες Web ή ιστοσελίδες επισκέπτονται σε ζευγάρια?
- **Κλασικά Παραδείγματα**
 - Στα ψιλκατζίδικα (που τα βρίσκει κανείς ανοικτά και τις Κυριακές) βρήκαν ότι συχνά οι πελάτες αγοράζουν μαζί Μπύρα και Πάνες.
 - Amazon.com: δείχνει συσχετιζόμενες αγορές
- **Στρατηγική Χρήσης τέτοιων στοιχείων**
 - Αποφάσισε αν πρέπει να βάλεις τα είδη κοντά (μαζί) για αύξηση του cross selling
 - Εναλλακτικά, τα βάζεις στην αρχή και στο τέλος του διαδρόμου ώστε οι πελάτες στη διαδρομή να πάρουν και άλλα προϊόντα

DBMS : Οι Προϊστορικοί Πρωταγωνιστές

- Αυτά τα Συστήματα παρουσιάστηκαν και έλαμψαν στην δεκαετία του 80
- Πολλές εγκαταστάσεις ΑΚΟΜΗ υπάρχουν – αλλά δεν γίνονται πωλήσεις και σιγά σιγά εξαφανίζονται...
 - IMS (IBM) -- Hierarchical Model (γλώσσα DL/1)
 - I-D-S (Honeywell) -- Network DBTG (Integrated Data Store)
 - IDMS (Cullinane) - Network (Integrated Data Mgmt System)
 - TOTAL (Cincom) - Network
 - IMAGE (Hewlett-Packard) - Network
 - SYSTEM 2000 (Intel-MRI) - Inverted (ad-hoc model)
 - Other Inverted: ADABAS (Software AG), Model 204 (CCA)
 - ...

DBMS : Παλιοί Πρωταγωνιστές

- **SYBASE**
- **INGRES** τώρα λέγεται, Computer Associates-Ask Group έγινε OPEN SOURCE σύστημα
- Άλλοι (Φθίνουν...):
 - **Rdb** (έκλεισε), **Gupta Quadbase**, **Ralma**, **Watcom**, **XDB**, ...

DBMS : Οι Πρωταγωνιστές

ΣΗΜΕΡΑ ΕΧΟΥΝ ΜΕΙΝΕΙ ΤΡΙΑ ΜΕΓΑΛΑ ΕΜΠΟΡΙΚΑ ΣΥΣΤΗΜΑΤΑ ΒΑΣΕΩΝ ΔΕΔΟΜΕΝΩΝ:

- **DB2**, σε κάθε IBM ή IBM-compatible πλατφόρμα (και όχι μόνο!, Unix, Linux, Windows, κλπ.)
- **ORACLE 12 (12C)**, σχεδόν παντού – πρώτη σε μερίδιο αγοράς – ειδικά σε Unix + μεγάλες εγκαταστάσεις
- **Microsoft SQL Server 2014**, σε Microsoft πλατφόρμες
 - **INFORMIX** (Αγοράστηκε από IBM!)

DBMS : Ειδικοί Πρωταγωνιστές

- Οι **MPP** πρωταγωνιστές (massively parallel):
 - **Terradata** (ο μεγαλύτερος), **Tandem** (NonStop SQL), **Oracle Parallel Server**, **Informix**, **Sybase** (Navigator), **DB2**, **DEC**,... (μερικοί έχουν κλείσει ήδη...)
- Οι μεγάλοι στον κόσμο των **Windows / PC**
 - **MICROSOFT ACCESS**
 - **Powersoft, Gupta...**
 - Χαρακτηριστικά: (a) SQL πρόσβαση (gateways)
 - (b) Εξάίρετα για Client-Server (DBMS)
 - (c) Μοιάζουν πολύ με τα «μεγάλα» DBMS

DBMS : Παλιοί Πρωταγωνιστές σε PC

- **Ακόμη στο PC**
 - **Paradox (Borland)**
 - **Q&A (Symantec)**
 - **FileMaker Pro (Claris Corp.)**
 - **DataEase Express**
 - **Approach (Lotus)**
 - **Alpha Four**
 - **Ακόμη Παλαιότερα: xBASE, dBASE, FoxPro, MicroRIM...**
- **Συνήθως έχουν ένα υποσύνολο των δυνατοτήτων των μεγάλων DBMS**

OODBMS Εταιρείες

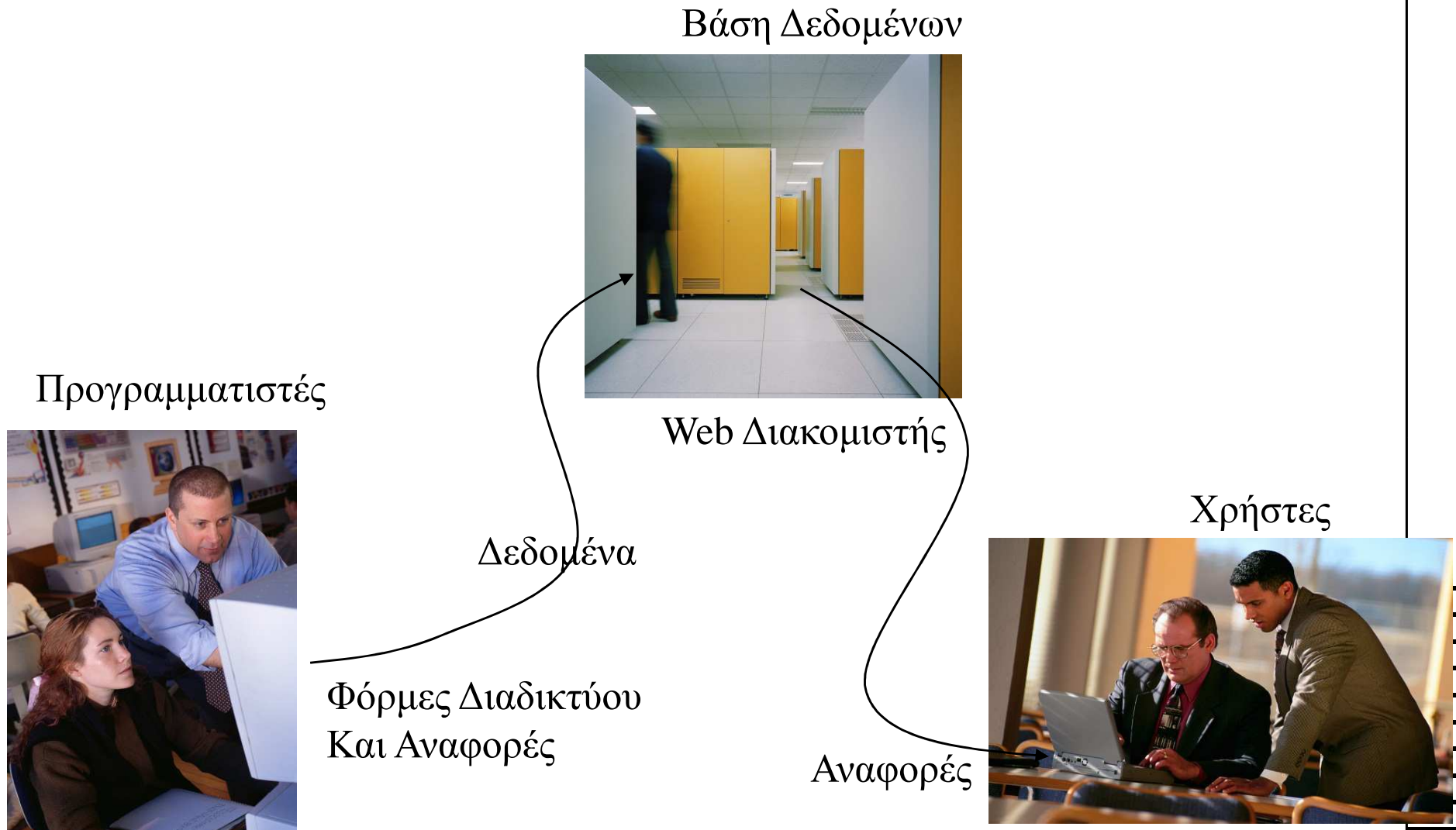
GemStone Systems, Inc.
Hewlett-Packard, Inc. (OpenODB)
IBEX Corporation, SA.
Illustra (Informix, Inc.)
Matisse Software, Inc.
O2 Technology, Inc.
Objectivity, Inc.
Object Design, Inc.
ONTOS, Inc.
POET Software Corporation
UniSQL
Unisys Corporation (OSMOS)
Versant Object Technology

OPEN DBMS

(Συστήματα Ανοικτού Κώδικα)

- MySQL
- PostgreSQL + EnterpriseDB
- Berkeley DB, Firebird, etc.

Web Βάσεις Δεδομένων



Web και Βάσεις Δεδομένων

- **Πάνω από 1 ΔΙΣ. HTML pages, 15 terabytes**
- **Τεράστιος Πλούτος Πληροφοριών**
 - Βιβλιοπωλεία, Εστιατόρια, Ταξιδιωτικό Υλικό, Αγορές, Λεξικά, Νέα, Χρηματαγορές, Οδηγοί, Χάρτες, ...
 - **Πολλών Μορφών**: κείμενα, εικόνες, φωνή, βίντεο...
 - **Πολλών Τύπων**: HTML, XML, postscript, pdf, JPEG, MPEG, MP3
- **Με Ιδιαίτερα Δυναμικά Χαρακτηριστικά**
 - Πάρα πολλές νέες σελίδες κάθε μέρα (> 1 εκατ.)
 - Δομή γράφου με συνδέσμους μεταξύ σελίδων (7-10 συνδέσμους)
- **Εκατοντάδες εκατομμυρίων ερωτήματα / πλοηγήσεις την ημέρα**

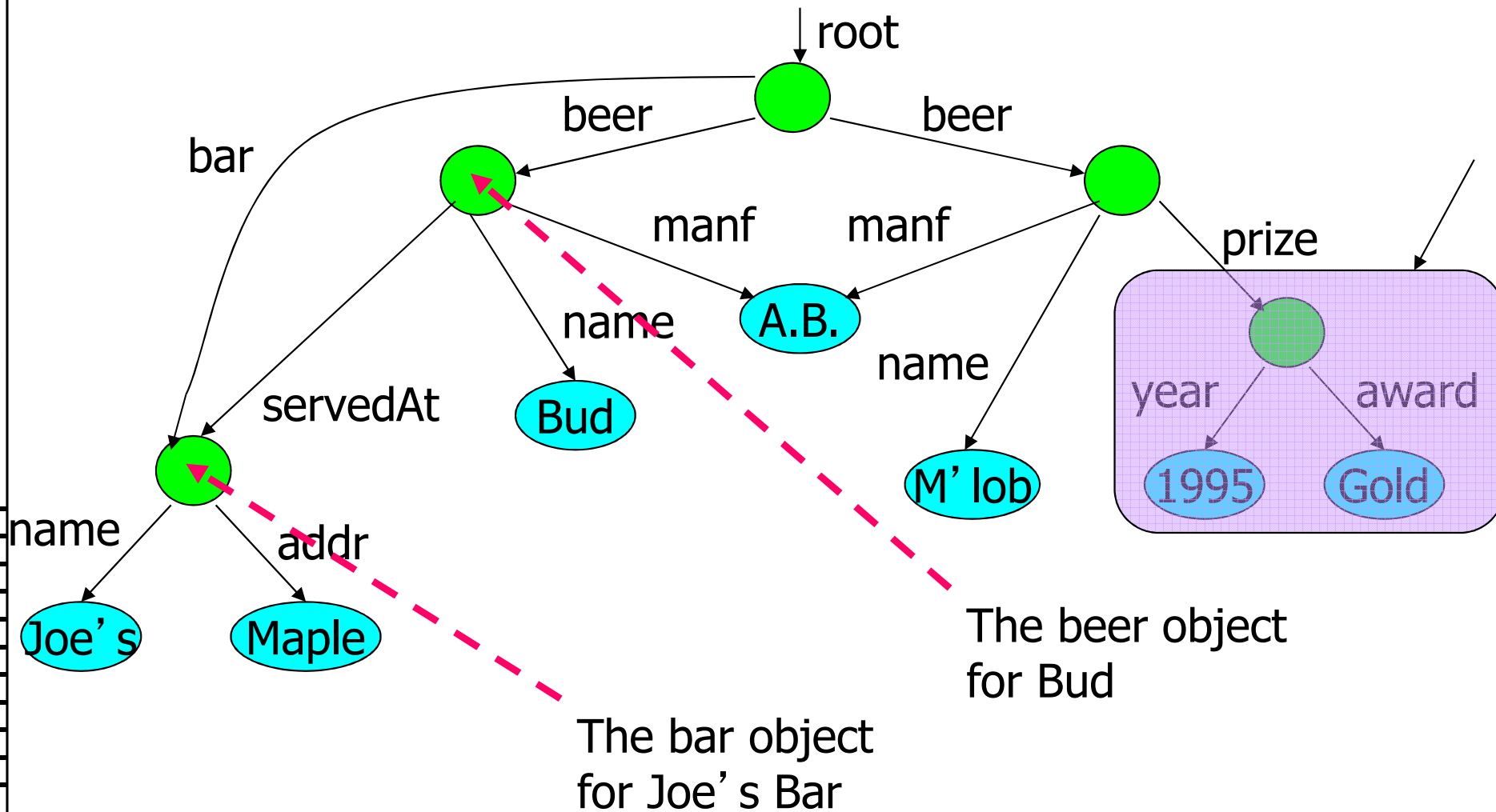
Το πρόβλημα Ολοκλήρωσης Πληροφοριών (Information Integration)

- Συσχετιζόμενα Δεδομένα υπάρχουν σε πολλούς (διαδικτυακούς) τόπους και θα μπορούσαν να δουλέψουν μαζί
- Όμως, οι διαφορετικές Βάσεις Δεδομένων έχουν πολλαπλές διαφορές:
 1. **Μοντέλο** (relational, object-oriented?).
 2. **Περιγραφή** (κανονικοποιημένο / μη-κανονικοποιημένο?).
 3. **Ορολογία**: είναι οι Σύμβουλοι και Υπάλληλοι? Οι υπεργολάβοι? Οι συνταξιούχοι?
 4. **Συμβατικοί Όροι** (μετρικό σύστημα, θερμοκρασία, κλπ)

Το Περίγραμμα

1. *Ολοκλήρωση Πληροφοριών: Το να κάνεις Βάσεις Δεδομένων από διάφορους τόπους να δουλεύουν σαν μια.*
2. *Ημι-δομημένα Δεδομένα (Semistructured Data) : Ένα νέο μοντέλο που έρχεται να αντιμετωπίσει τα προβλήματα της ολοκλήρωσης πληροφοριών*
3. *XML : Μια ειδική γλώσσα για την περιγραφή των semistructured data schemas και για την παράσταση των δεδομένων.*

Γράφος για Semistructured Data



XML

- XML = Extensible Markup Language.
- Ενώ η HTML χρησιμοποιεί ετικέτες για μορφοποίηση (π.χ., “*italic*”), η XML χρησιμοποιεί ετικέτες (tags) για σημασιολογία - semantics (π.χ., “αυτό είναι μια διεύθυνση”).
- Βασική Ιδέα: Δημιούργησε σύνολα ετικετών για ένα τομέα (π.χ., γονιδιωματική), και μετάφρασε όλα τα δεδομένα σε XML έγγραφα.

Παράδειγμα: Well-Formed XML

<? XML VERSION = "1.0" STANDALONE = "yes" ?>

<BARS>

<BAR><NAME>Joe's Bar</NAME>

<BEER><NAME>Bud</NAME>

<PRICE>2.50</PRICE></BEER>

<BEER><NAME>Miller</NAME>

<PRICE>3.00</PRICE></BEER>

</BAR>

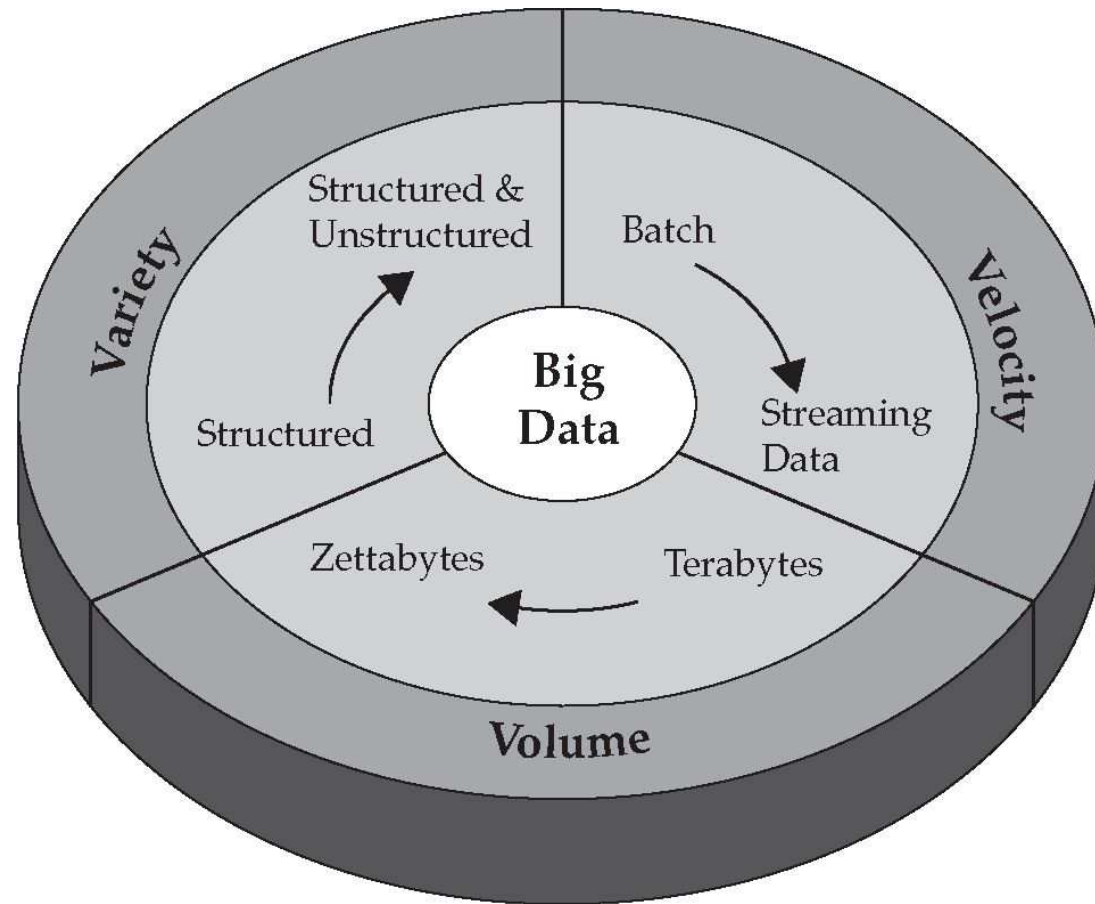
<BAR> ...

</BARS>

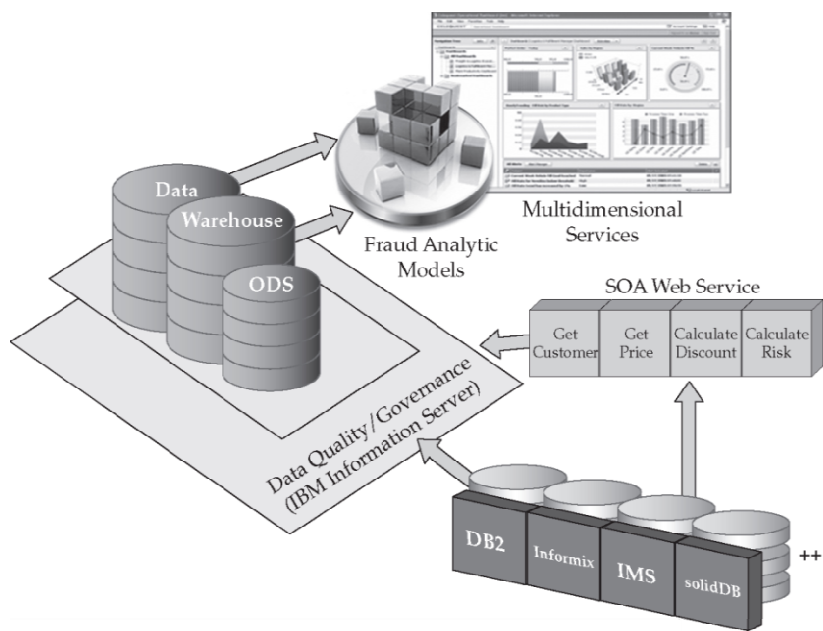
BIG DATA - Χαρακτηριστικά

- **ΤΕΡΑΣΤΙΟΣ ΟΓΚΟΣ** -- Το 2000, είχαμε περίπου 800,000 petabytes (PB) αποθηκευμένα δεδομένα σε όλο τον κόσμο. Περιμένουμε το νούμερο να φτάσει στα 35 zettabytes (ZB) μέχρι το 2020. (Μόνο το Twitter παράγει περισσότερο από 7 terabytes (TB) δεδομένων κάθε μέρα, το Facebook 10 TB, και μερικές επιχειρήσεις πολλά terabytes κάθε ώρα.
- **ΠΟΙΚΙΛΙΑ** -- Δομημένα, Ημί-δομημένα, Αδόμητα
- **ΤΑΧΥΤΗΤΑ** παραγωγής (Streams)

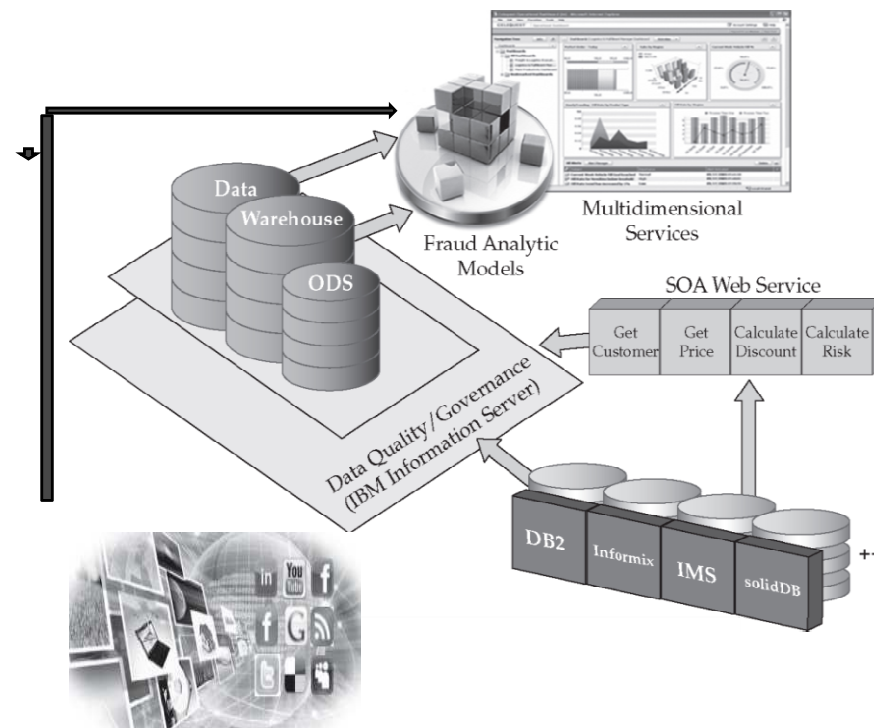
BIG DATA - Χαρακτηριστικά



Χρήση του 80% δεδομένων (μη-σχεσιακά)



Κλασικό Μοντέλο



Νέο Μοντέλο

HADOOP

- Hadoop (<http://hadoop.apache.org/>) είναι ένα σημαντικό Apache project στο Apache Software Foundation που είναι γραμμένο στη Java.
- Hadoop είναι ένα ΥΠΟΛΟΓΙΣΤΙΚΟ ΠΕΡΙΒΑΛΛΟΝ κτισμένο πάνω σε ένα κατανεμημένο σύστημα αρχείων που φτιάχτηκε ειδικά για πολύ υψηλής τάξης μεγέθους επεξεργασία δεδομένων.
- Hadoop έχει εμπνευστεί από τη δουλειά της Google στο δικό της σύστημα αρχείων (GFS) και το προγραμματιστικό περιβάλλον MapReduce, όπου η δουλειά σπάει σε Mapper και Reducer εργασίες για την επεξεργασία δεδομένων αποθηκευμένων σε μια συστάδα από servers για παραλληλισμό.

Χαρακτηριστικά των NoSQL Συστημάτων

- Δεν χρησιμοποιούν SQL
- Δεν χρησιμοποιούν Πίνακες (relations)
- Δεν βασίζονται σε ΣΥΣΧΕΤΙΣΕΙΣ (relationships)
- Δεν έχουν την ιδιότητα ACID

- ΔΕΝ ΕΧΟΥΝ ΤΥΠΙΚΟ ΣΧΗΜΑ (formal schema)
- Προσαρμοσμένα σε WEB ανάπτυξη εφαρμογών
- Προσαρμοσμένα σε ανάπτυξη ΜΕΓΑΛΩΝ ΕΦΑΡΜΟΓΩΝ
- Open Source

Είδη NoSQL Συστημάτων (a)

■ DOCUMENT STORES

– Παραδείγματα: CouchDB, MongoDB

```
{  
  "title" : "Foundations of Databases"  
  "rating" : 10  
}
```

Είδη NoSQL Συστημάτων (b)

- **KEY – VALUE Stores**

- Όλα αποθηκεύονται ως Ζεύγη (Key, Value)

- Παραδείγματα: Memcached, Riak, ...

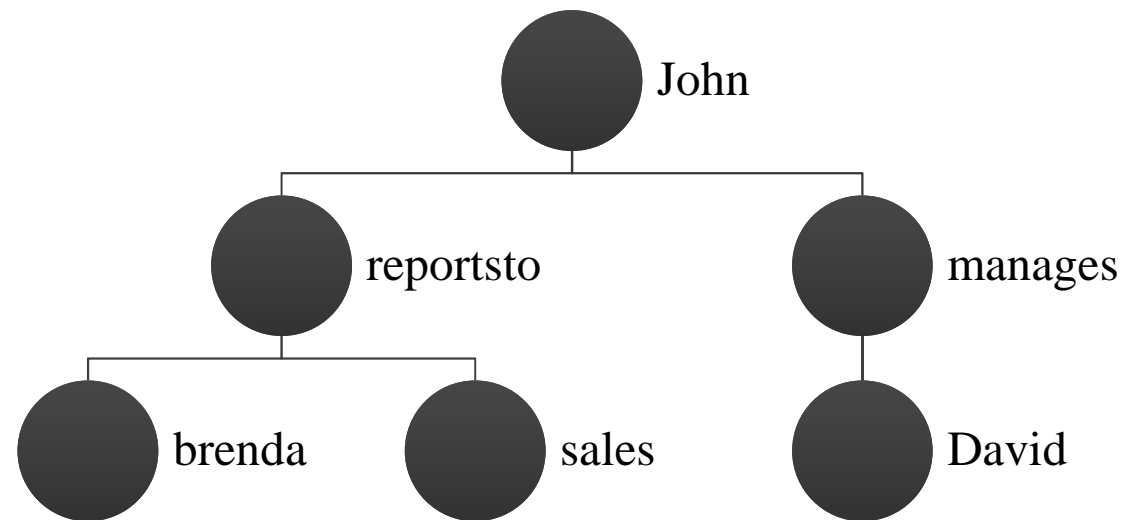
KEY	VALUE
Name1	bob
Course_44	DBMS
V32_phote	{binary data}
...	

Είδη NoSQL Συστημάτων (c)

■ GRAPH DATABASES

– Όλα αποθηκεύονται ως Γράφοι

- Παραδείγματα: Neo4j, AlegroGraph, DB2 NoSQL, ...



Γιατί NoSQL???

- *Αν θέλεις* **ΕΥΕΛΙΚΤΟ ΣΧΗΜΑ**
- *Αν έχεις* **τεράστιο όγκο από δεδομένα**
- *Αν δεν σε ενδιαφέρει πολύ η* **ΣΥΝΕΠΕΙΑ** στα δεδομένα

Υπολογιστική Νέφους με Βάσεις Δεδομένων

Cloud Computing with Databases

Google: BigTable, για εσωτερική αποθήκευση
AppEngine: <http://code.google.com/appengine/>
Άριστο για σύνθετα έγγραφα / αντικείμενα
ΔΕΝ είναι SQL; ΔΕΝ γίνονται συνενώσεις (JOIN)

Generic: Hadoop (Apache) - Λογισμικό Ανοικτού Κώδικα για Cloud

Amazon:

S3	Αρχεία, ιδιαίτερα μεγάλα http://aws.amazon.com/s3/
SimpleDB	Παρεμφερές με το BigTable http://aws.amazon.com/simpledb/
RDS	Σχεσιακού Μοντέλου Υπηρεσία MySQL ή Oracle 12g http://aws.amazon.com/rds/

Microsoft:

Azure	SQL Server http://www.microsoft.com/windowsazure/
-------	---

Πλεονεκτήματα της Χρήσης Υπηρεσιών Νέφους για DBMS

- **Δεν υπάρχουν σταθερά κόστη**
 - Δεν απαιτείται Υλικό ή Λογισμικό
 - Δεν απαιτείται συντήρηση
 - Εύκολη Διαχείριση
- **Το Τίμημα καθορίζεται ανάλογα με τη χρήση**
 - Μηνιαίο Κόστος ανάλογα με το μέγεθος της ΒΔ
 - Μηνιαίο Κόστος ανάλογα με τη χρήση (data transfer)
- **Επεκτασιμότητα**
 - Πολλαπλοί, κατανεμημένοι διακομιστές / εξυπηρετητές
 - Πολλαπλές, υψηλής ταχύτητας Internet συνδέσεις
- **Αξιοπιστία**
 - Κατανεμημένο
 - Σύστημα που το τρέχουν ειδικοί
 - Ελεγχόμενη ασφάλεια

Όρια με τις Cloud Βάσεις Δεδομένων

- Τα κόστη καμιά φορά ξεφεύγουν – ιδιαίτερα όταν η χρήση είναι πολύ συχνή
- Τότε ίσως είναι φτηνότερα να πάει κανείς στη λύση αγοράς και κλασσικής χρήσης (in-house)

Τμήμα - Cloud Database Pricing

Παράδειγμα: Amazon RDS (MySQL), U.S. East

1 Extra large instance
20 hours/day
20 GB/month at 50 million I/O per month
10 GB/month data transfer in
500 GB/month data transfer out
20 GB/month regional transfer
=> \$616 per month (\$7400/year)

Παράδειγμα : Microsoft SQL Azure Business Edition

1 Extra large instance (\$0.96/hour = \$576/month)
20 GB/month (\$200/month)
10 GB/month data transfer in (\$1/month)
500 GB/month data transfer out (\$75/month)
=> \$852 per month (\$10,224/year)

Cloud and Databases



“It was much nicer before people started ^{misplacing} storing all their data in the Cloud.”

ΕΞΕΛΙΞΗ στην Εφαρμογή και Τεχνολογία Διοίκησης / Διαχείρισης Δεδομένων

Απλή Διαχείριση
Δεδομένων

Επιχειρησιακή
Διαχείριση - ΟΠΣ

Αρχές 80

Τέλη 80

Αρχές - Μέσα 90

Τέλη 90 – 21ος Αιώνας

Προ- Σχεσιακά
DBMS

Πρώτα
Σχεσιακά

Client-server
Σχεσιακά

Επιχειρησιακή
Κάλυψη

Internet
Computing

Απλό
OLTP

Ενεργές
Βάσεις
Δεδομένων

Αποθήκες
Δεδομένων &
Hi-end OLTP

Πακέτα & Κάθετες
Εφαρμογές -
Business Intelligence

Simple
transactions,
on-line
backup &
recovery

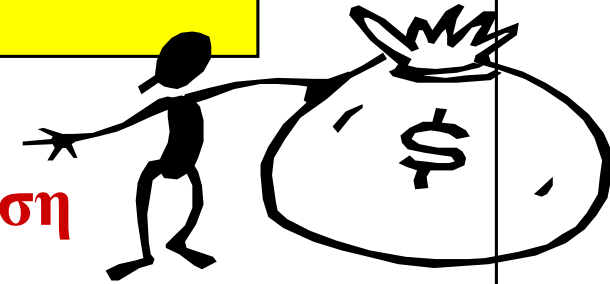
Stored
procedures,
triggers

Scaleable OLTP,
parallel query,
partitioning,
cluster support,
row-level locking,
high availability

Support for all
types of data,
extensibility,
objects

Middleware
(messaging,
queues, events)
Java, CORBA,
Web interfaces
XML)

ΤΕΛΙΚΕΣ ΣΚΕΨΕΙΣ



- Το DBMS χρησιμοποιείται για την συντήρηση ΜΕΓΑΛΩΝ Βάσεων Δεδομένων.
- Μεταξύ των πλεονεκτημάτων, είναι η Ανάκαμψη από Καταστροφές του συστήματος, η Συνδρομικότητα, γρήγορη ανάπτυξη εφαρμογών, ακεραιότητα, ασφάλεια.
- Πεδία αφαίρεσης βοηθούν στην ανεξαρτησία των προγραμμάτων από τις Φυσικές Δομές.
- Οι Διαχειριστές (DBA) έχουν υπεύθυνες εργασίες και πληρώνονται πολύ καλά!
- DBMS R&D είναι από τις πλέον ελκυστικές περιοχές της Πληροφορικής (τέλειος συνδυασμός Πράξης – Θεωρίας)