



**EPL646 – Advanced Topics in
Databases**

Lecture 1

**Syllabus and Course
Overview**

Demetris Zeinalipour

<http://www.cs.ucy.ac.cy/~dzeina/courses/epl646>



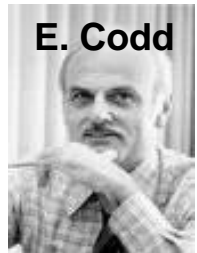
Στόχοι ΕΠΛ646;

- **Στόχοι:**
 - **Κατανόηση και Υλοποίηση** προχωρημένων εννοιών που αφορούν την **εσωτερική λειτουργία** μιας σχεσιακής βάσης δεδομένων
 - Έκθεση σε **Προχωρημένα και Ανερχόμενα Θέματα** στο πεδίο των βάσεων δεδομένων (web, cloud, sensor, spatio-temporal, indoor, κτλ.)
 - Να επιτρέψει στους φοιτητές να αποκτήσουν ένα **ισχυρό υπόβαθρο** στις Βάσεις Δεδομένων καθιστώντας τους ικανούς να **αξιοποιήσουν** τις γνώσεις τους σε άλλα πεδία της Πληροφορικής.

ΕΠΛ646: Εισαγωγή (Χθές, Σήμερα, Αύριο)



- **Βάση Δεδομένων (Database):** Συλλογή από *ενοποιημένα* – *integrated* - δεδομένα).
- **DBMS (Database Management System)**
 - Ένα λογισμικό πακέτο το οποίο έχει σχεδιαστεί για να **αποθηκεύει** και να **διαχειρίζεται** βάσεις δεδομένων
- **R(elational)DBMS:** Σχεσιακή DBMS (δεδομένα αναπαριστώνται στο **σχεσιακό μοντέλο**)
 - Σε αυτό το μοντέλο, τα δεδομένα αναπαριστώνται σε **πίνακες** + **περιορισμοί** που διασφαλίζονται από το DBMS.
 - Το μοντέλο προκάλεσε μια **επανάσταση** στο χώρο των βάσεων δεδομένων λόγω της **απλότητας** και του **μαθηματικού** του **υπόβαθρου**:
 - **1969:** Το Σχεσιακό Μοντέλο υλοποιείται από τη βάση **IBM System R**
 - **1970:** Η IBM δημιουργεί την **SEQUEL** (προπομπό της **SQL**)
 - **1981:** Ο **Codd** παίρνει το **Turing Award** στη πληροφορική
 - **1985:** Η **IBM** κάνει την **SQL** Πατέντα (US Pat. 4,506,326).
 - **Σήμερα:** Το Σχεσιακό Μοντέλο υλοποιείται από τις περισσότερες σύγχρονες βάσεις δεδομένων (Oracle, IBM DB2, SQL Server, PostgreSQL, MySQL, κτλ) αποτελώντας το υπόβαθρο των επιχειρήσεων (enterprise environments)

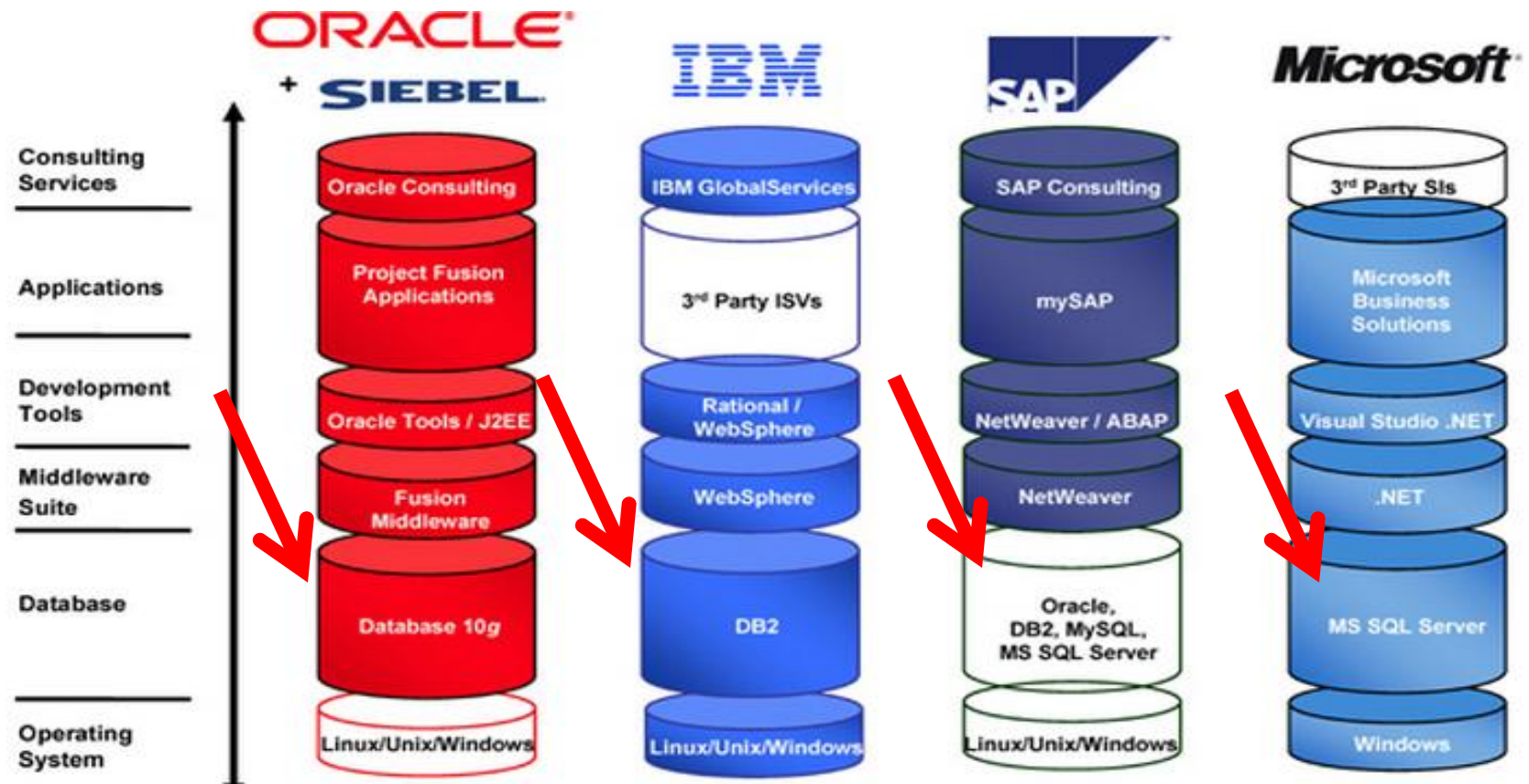


E. Codd

ΕΠΛ646: Εισαγωγή (Χθές, Σήμερα, Αύριο)



RDBMS ως υπόβαθρο των Επιχειρηματικών Εφαρμογών (Enterprise Applications)



ΕΠΛ646: Εισαγωγή (Χθές, Σήμερα, Αύριο)

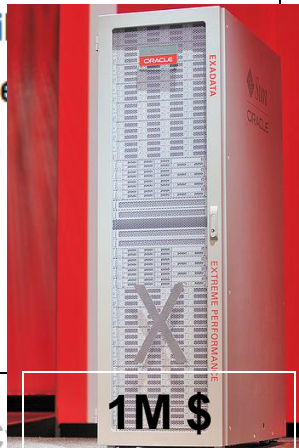


Larry Ellison

ORACLE

The Information Company (37.1B\$ / '12)

Server and Storage Systems	Database	Middleware	Applications	Industries
<div>Sun Servers</div> <div>Storage and Tape</div> <div>Exadata Database Machine</div> <div>SPARC SuperCluster T4-4</div> <div>Database Appliance</div> <div>Exalogic Elastic Cloud</div> <div>Oracle Solaris</div> <div>Oracle Linux</div> <div>Virtualization</div> <div>Enterprise Manager Ops Center</div> <div>More Servers and Storage ▾</div>	<div>Oracle Database 11g</div> <div>Real Application Clusters</div> <div>Data Warehousing</div> <div>Database Security</div> <div>Exadata Database Machine</div> <div>Database Appliance</div> <div>Big Data</div> <div>Enterprise Manager for Database</div> <div>Embedded</div> <div>High Availability</div> <div>MySQL</div> <div>More Database ▾</div>	<div>Java</div> <div>WebLogic Server</div> <div>Exalogic Elastic Cloud</div> <div>Exalytics In-Memory Machine</div> <div>SOA BPM</div> <div>Social Network</div> <div>WebCenter</div> <div>Content Portal</div> <div>Business Analytics</div> <div>Identity Management</div> <div>Enterprise Manager for Middleware</div> <div>Data Integration</div> <div>More Middleware ▾</div>	<div>Oracle Fusion Applications</div> <div>Oracle E-Business Suite</div> <div>Human Capital Management</div> <div>PeopleSoft (HRM)</div> <div>RightNow</div> <div>Siebel (CRM)</div> <div>ATG</div> <div>Oracle CRM On Demand</div> <div>JD Edwards EnterpriseOne</div> <div>JD Edwards World</div> <div>Hyperion</div> <div>Primavera</div> <div>Application Integration</div> <div>More Applications ▾</div>	<div>Communications</div> <div>Financial Services</div> <div>Healthcare</div> <div>High Technology</div> <div>Insurance</div> <div>Life Sciences</div> <div>Public Sector</div> <div>Retail</div> <div>Utilities</div> <div>More</div>



Τι καλύπτει το ΕΠΛ646;

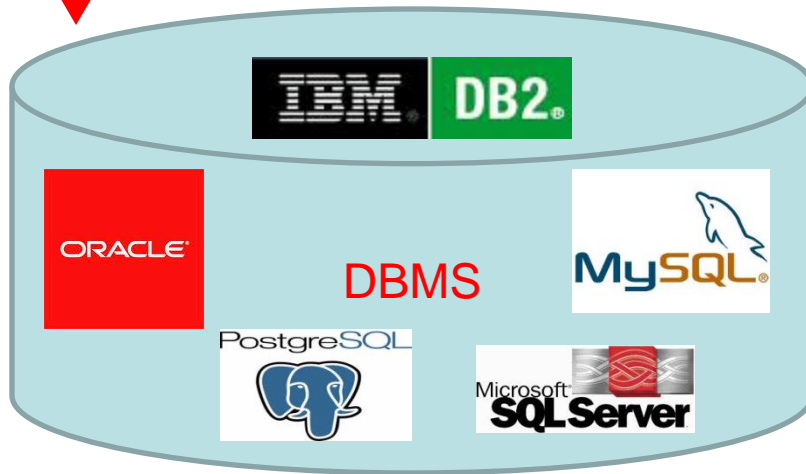


EPL342 – DBs (Modeling, SQL, Normalization)

Programmers / Users

SQL

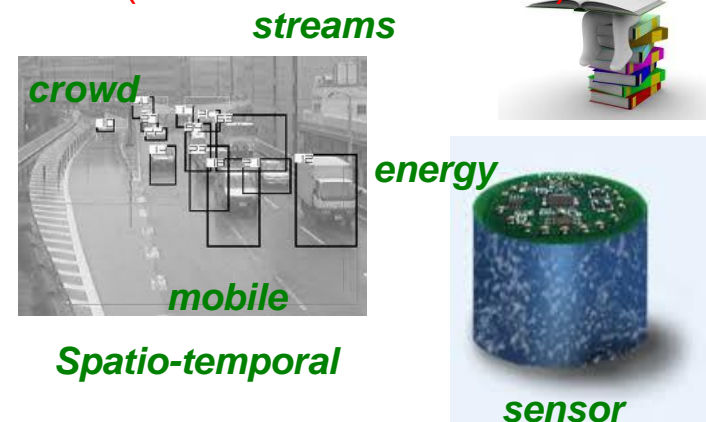
EPL646 - PART A (RDBMS Internals)



EPL646 - PART B (Distributed/Web/Cloud DBs)



EPL646 - PART C (Other DB Research)





Συμβόλαιο Μαθήματος

- **Επίπεδο:** Μεταπτυχιακό
 - Επιλογή για όλες τις Κατευθύνσεις
- **Πίστωση:** 8 μονάδες ECTS
- **Προαπαιτούμενα:**
 - ΕΠΛ342: Βάσεις Δεδομένων (ή αντίστοιχο) - (ER Modeling, SQL, DB Programming, Normalization)
- **Μέθοδοι Διδασκαλίας**
 - Διαλέξεις (3 ώρες εβδομαδιαίως)
 - Φροντιστήριο/Εργαστήριο (1 1/2 ώρα εβδομαδιαίως)
- **Υπόβαθρο**
 - Επαρκή γνώση σε συστήματα Linux (ΕΠΛ371) και προγραμματισμός σε γλώσσες C/C++/JAVA (ΕΠΛ132)



Συμβόλαιο Μαθήματος

- **Αξιολόγηση**

- 50% Τελική Εξέταση (1)

- 20% Ενδιάμεση Εξέταση (1)

- Προκαταρτική Ημερομηνία:
22/10/14 (7^η βδομάδα)!

- 30% Ασκήσεις

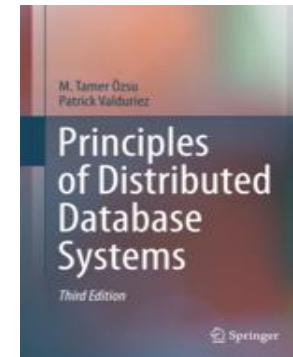
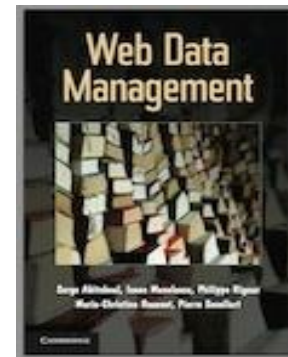
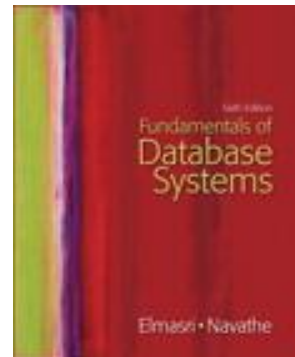
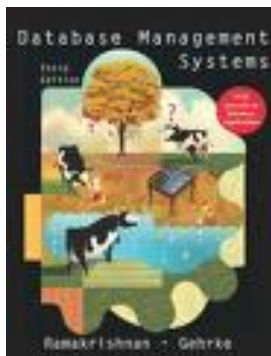
- Προγραμματιστικές/Θεωρητικές Ασκήσεις

- Παρουσίαση Μελέτης / Άρθρων Εξαμήνου



Βιβλιογραφία

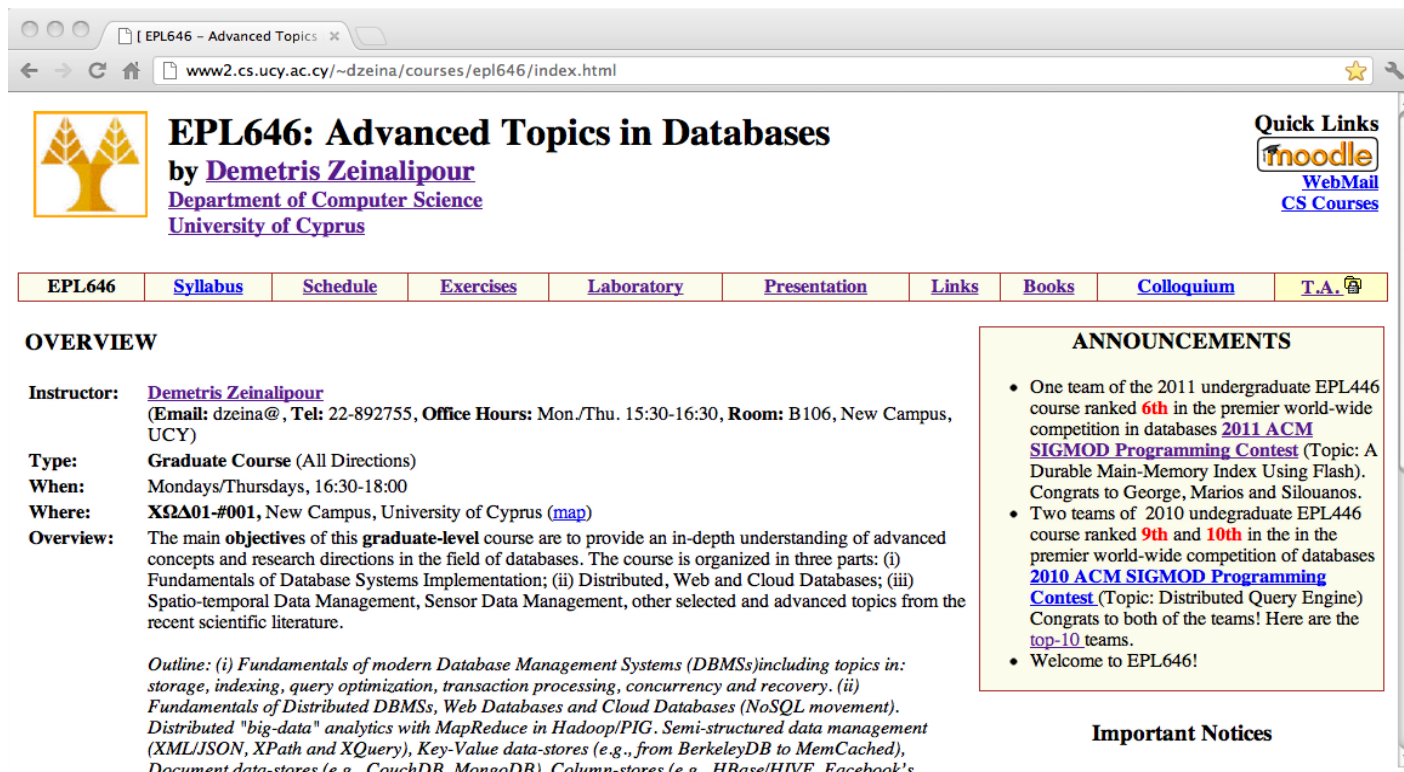
- Σημειώσεις Μαθήματος και Επιλεγμένη Αρθρογραφία
- **Fundamentals of Database Systems**, 6/E Ramez Elmasri, Shamkant B. Navathe ISBN: ISBN-10: 0136086209, Addison-Wesley, 2011, 1200 pp, 03/30/2010
- **Web Data Management**, Serge Abiteboul, Ioana Manolescu, Philippe Rigaux, Marie-Christine Rousset, Pierre Senellart; ISBN-10: 1107012430, ISBN-13: 978-110701243, Cambridge University Press, 450 pages, (available online), 2011.
- **Principles of Distributed Database Systems**, Özsu, M. Tamer, Valduriez, Patrick, 3rd Edition, 846 p., Springer Press, 2011.
- **Database Management Systems**, 3rd Edition Ramakrishnan, & Johannes Gehrke, 1104 pp. McGraw-Hill Publisher, ISBN 0-07-123057-2, 2003.





Πληροφορίες σχετικά με το μάθημα:

<http://www.cs.ucy.ac.cy/~dzeina/courses/epl646>



The screenshot shows a web browser window with the address bar displaying `www2.cs.ucy.ac.cy/~dzeina/courses/epl646/index.html`. The page title is "EPL646: Advanced Topics in Databases" by Demetris Zeinalipour, Department of Computer Science, University of Cyprus. The page features a navigation menu with links: EPL646, Syllabus, Schedule, Exercises, Laboratory, Presentation, Links, Books, Colloquium, and T.A. There is also a "Quick Links" section with links to Moodle, WebMail, and CS Courses. The main content area is divided into two sections: "OVERVIEW" and "ANNOUNCEMENTS".

EPL646: Advanced Topics in Databases
by **Demetris Zeinalipour**
Department of Computer Science
University of Cyprus

Quick Links
Moodle
WebMail
CS Courses

EPL646	Syllabus	Schedule	Exercises	Laboratory	Presentation	Links	Books	Colloquium	T.A.
--------	----------	----------	-----------	------------	--------------	-------	-------	------------	------

OVERVIEW

Instructor: **Demetris Zeinalipour**
(Email: dzeina@, Tel: 22-892755, Office Hours: Mon./Thu. 15:30-16:30, Room: B106, New Campus, UCY)

Type: Graduate Course (All Directions)

When: Mondays/Thursdays, 16:30-18:00

Where: XΩΔ01-#001, New Campus, University of Cyprus (map)

Overview: The main objectives of this graduate-level course are to provide an in-depth understanding of advanced concepts and research directions in the field of databases. The course is organized in three parts: (i) Fundamentals of Database Systems Implementation; (ii) Distributed, Web and Cloud Databases; (iii) Spatio-temporal Data Management, Sensor Data Management, other selected and advanced topics from the recent scientific literature.

Outline: (i) Fundamentals of modern Database Management Systems (DBMSs) including topics in: storage, indexing, query optimization, transaction processing, concurrency and recovery. (ii) Fundamentals of Distributed DBMSs, Web Databases and Cloud Databases (NoSQL movement). Distributed "big-data" analytics with MapReduce in Hadoop/PIG. Semi-structured data management (XML/JSON, XPath and XQuery), Key-Value data-stores (e.g., from BerkeleyDB to MemCached), Document data-stores (e.g., CouchDB, MongoDB), Column-stores (e.g., HBase/HIVE, Facebook's

ANNOUNCEMENTS

- One team of the 2011 undergraduate EPL446 course ranked **6th** in the premier world-wide competition in databases **2011 ACM SIGMOD Programming Contest** (Topic: A Durable Main-Memory Index Using Flash). Congrats to George, Marios and Silouanos.
- Two teams of 2010 undergraduate EPL446 course ranked **9th** and **10th** in the in the premier world-wide competition of databases **2010 ACM SIGMOD Programming Contest** (Topic: Distributed Query Engine) Congrats to both of the teams! Here are the **top-10** teams.
- Welcome to EPL646!

Important Notices



WWW

- Για τις εκπαιδευτικές δραστηριότητες του μαθήματος (υποβολή εργασιών, φόρουμ ανακοινώσεων, ερωτηματολόγια, βαθμολογίες εργασιών, κτλ) θα χρησιμοποιηθεί το Moodle. <http://moodle.cs.ucy.ac.cy/>

The screenshot displays the Moodle LMS interface for the course **EPL646: Advanced Topics in Databases** (ΕΠΛ646: Προχωρημένα Θέματα Βάσεων Δεδομένων). The user is logged in as **Demetris Zeinalipour**. The interface includes a sidebar with navigation links (People, Activities, Search Forums, Administration), a main content area with the course title and instructor information, and a right sidebar with news, events, and activity sections.

People

- Participants

Activities

- Forums

Search Forums

Advanced search

Administration

- Turn editing off
- Settings
- Assign roles
- Grades
- Groups
- Backup
- Restore
- Import

Topic outline

EPL646: Advanced Topics in Databases

(ΕΠΛ646: Προχωρημένα Θέματα Βάσεων Δεδομένων)

Instructor: Demetris Zeinalipour

For more information on this course please visit <http://www2.cs.ucy.ac.cy/~dzeina/courses/epl646/>

Important Notices:

- Please check this section regularly for additions and changes.
- To view all resources available on this page you need to logon.

Latest News

Add a new topic...

(No news has been posted yet)

Upcoming Events

There are no upcoming events

Go to calendar...
New Event...

Recent Activity

Activity since Wednesday, 29 August 2012, 01:25 PM

Full report of recent activity...

Nothing new since your last login

Blocks

ΕΠΛ646: Ενότητα Α

Εσωτερική Λειτουργία ενός RDBMS

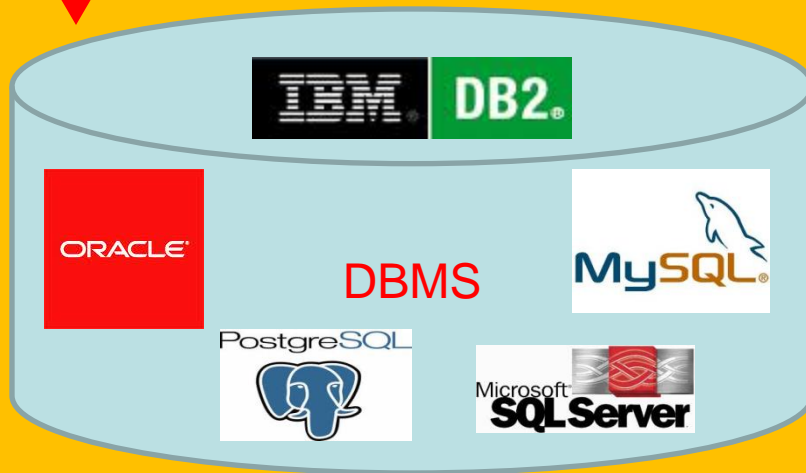


EPL342 – DBs (Modeling, SQL, Normalization)

Programmers / Users

SQL

EPL646 - PART A
(RDBMS Internals)

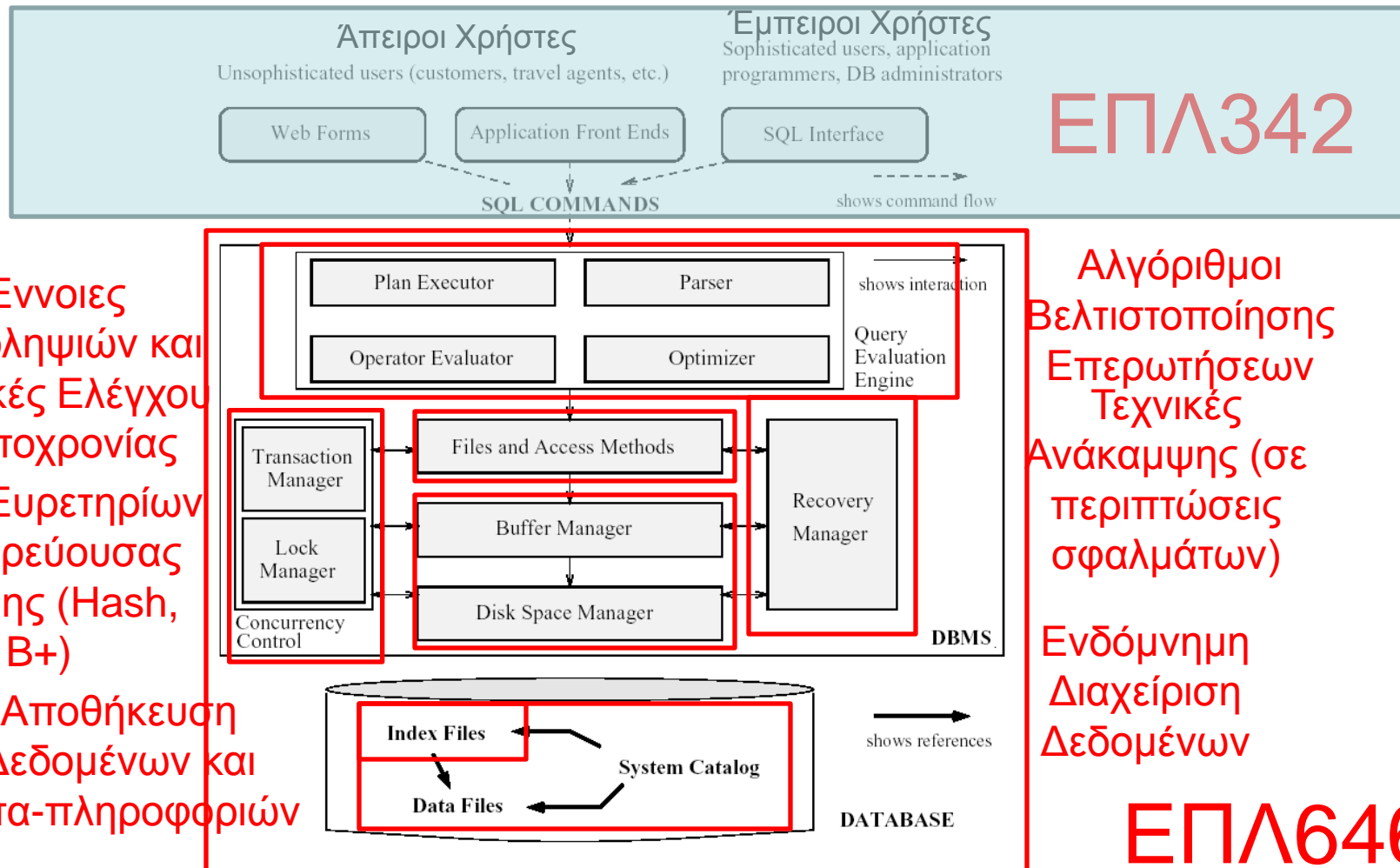
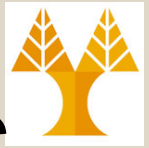


EPL646 - PART B
(Distributed/Web/Cloud DBs)

EPL646 - PART C
(Other DB Research)

ΕΠΛ646: Ενότητα Α

Εσωτερική Λειτουργία ενός RDBMS



Έννοιες
Δοσοληψιών και
Τεχνικές Ελέγχου
Ταυτοχρονίας
Δομή Ευρετηρίων
Δευτερεύουσας
Μνήμης (Hash,
B+)
Αποθήκευση
Δεδομένων και
Μετα-πληροφοριών

Αλγόριθμοι
Βελτιστοποίησης
Επερωτήσεων
Τεχνικές
Ανάκαμψης (σε
περιπτώσεις
σφαλμάτων)

Ενδόμνημη
Διαχείριση
Δεδομένων

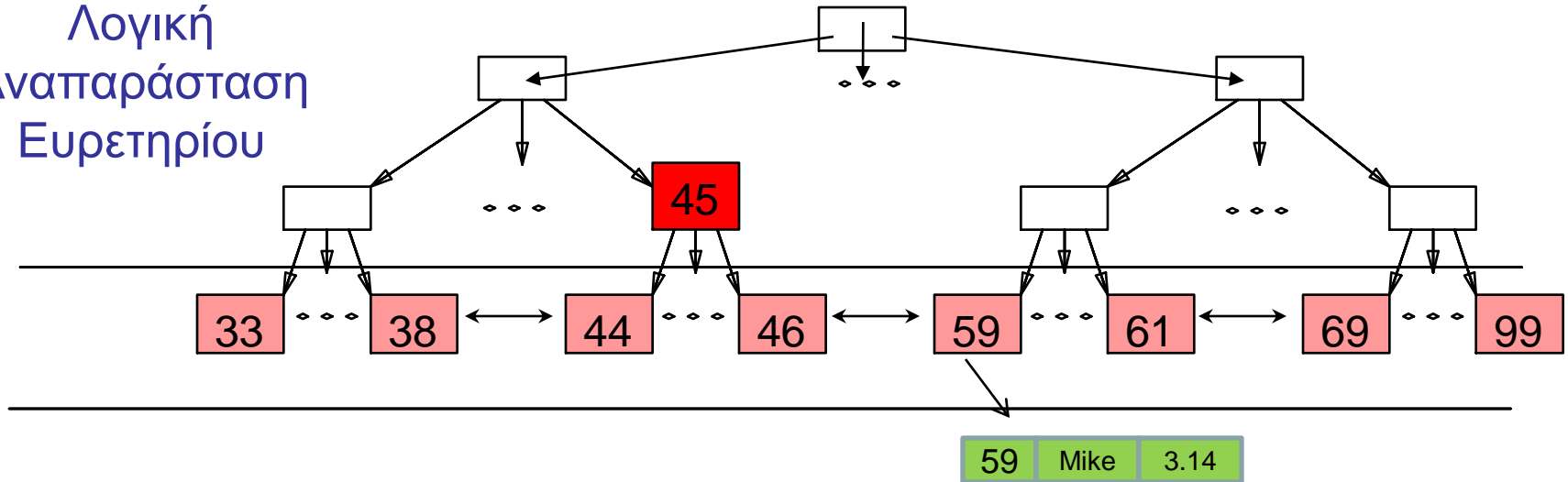
ΕΠΛ646

ΕΠΛ646: Ενότητα Α

((Disk-based) Index Structures)

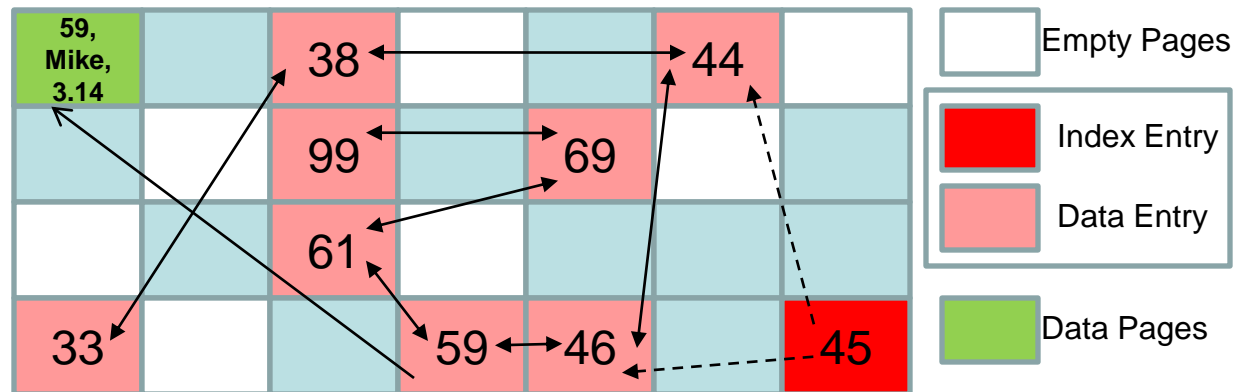


Λογική
Αναπαράσταση
Ευρετηρίου



Φυσική
Αναπαράσταση
Ευρετηρίου στη
Δευτερεύουσα
Μνήμη

Physical Layout (on Disk)



ΕΠΛ646: Ενότητα Α

(Βελτιστοποίηση Επερωτήσεων)



Αναλυτής (Parser): Αναλύει τα SQL επερωτήματα του χρήστη και τα μεταφέρει στον Βελτιστοποιητή

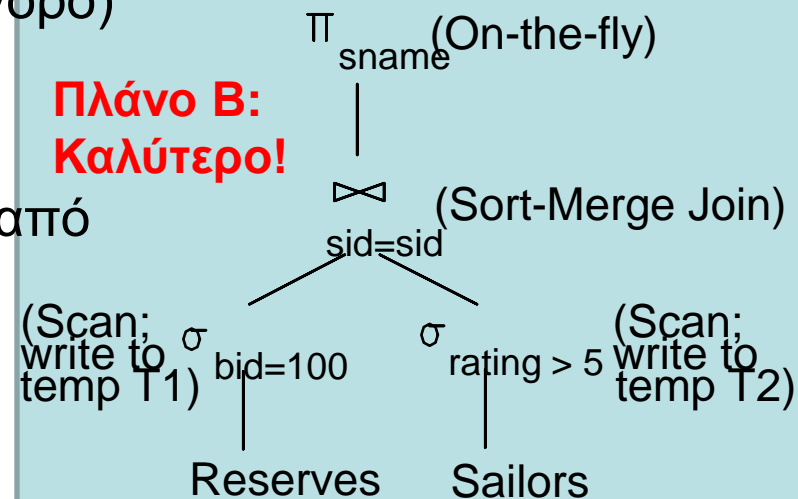
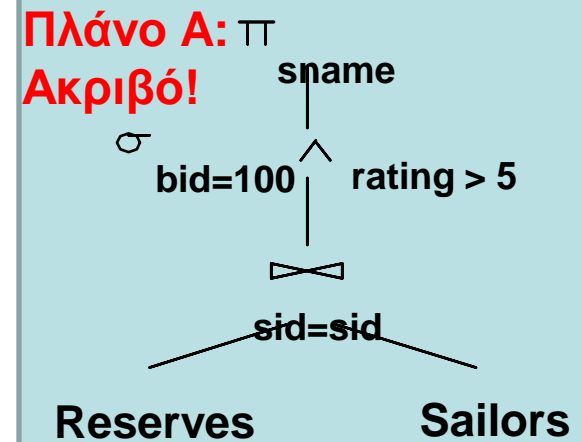
Βελτιστοποιητής (Optimizer): Κάνει χρήση μέτα-πληροφοριών στον **κατάλογο συστήματος (system catalog)** για να γνωρίζει τα διαθέσιμα ευρετήρια, τον αριθμό των πλειάδων σε ένα πίνακα.

Όλα αυτά για να βρει το καλύτερο (γρήγορο) πλάνο εκτέλεσης!

Εκτελεστής Πλάνου (Plan Executor):

Εύρεση και εκτέλεση φθηνότερου πλάνου από όλα τα δένδρα σχεσιακών τελεστών.

```
SELECT S.sname
FROM Reserves R, Sailors S
WHERE R.sid=S.sid AND
      R.bid=100 AND S.rating>5
```



ΕΠΛ646: Ενότητα Α

(Δοσοληψίες)



- Δοσοληψία (transaction), μια ατομική (*atomic, δηλ. all-or-nothing*) ακολουθία από read / write στη βάση.
 - Transaction Example in MySQL
START TRANSACTION;
SELECT @A:=SUM(salary) FROM table1 WHERE type=1;
UPDATE table2 SET summary=@A WHERE type=1;
UPDATE table3 SET summary=@A WHERE type=1;
COMMIT;
- Κάθε δοσοληψία, που ολοκληρώνεται, πρέπει να αφήνει την DB σε **συνεπή κατάσταση** (consistent state).
 - Οι κανόνες ακεραιότητας (integrity constraints), π.χ., Primary Key, Foreign Key, Check, Not Null, Unique, επιβάλλονται αυτόματα από μια βάση.
 - Από εκεί και πέρα, η RDBMS δεν γνωρίζει τους **επιχειρησιακούς** κανόνες ακεραιότητας (που ορίζονται μέσω των δοσοληψιών)

ΕΠΛ646: Ενότητα Α

(Έλεγχος Ταυτοχρονίας)



- Η **παράλληλη εκτέλεση** των δοσοληψιών είναι απαραίτητη για να έχει ένα DBMS **καλή επίδοση**
 - Αυτό διότι η **πρόσβαση στη δευτερεύουσα μνήμη (δίσκο)** είναι συχνή, και **σχετικά αργή**, συνεπώς είναι σημαντικό να κρατάμε τον επεξεργαστή απασχολημένο!
- **Παρεμβάλλοντας (Interleaving)** τις δοσοληψιών μπορεί να προκαλέσει **ασυνέπεια (inconsistency)**: π.χ., μια επιταγή αποπληρώνεται ενώ υπολογίζεται το ισοζύγιο του λογαριασμού.... το αποτέλεσμα του ισοζυγίου είναι λανθασμένο!
- Το DBMS διασφαλίζει ότι τέτοια προβλήματα δε θα προκύψουν: *Οι χρήστες έχουν την εντύπωση ότι οι δοσοληψίες τους εκτελούνται σειριακά!*

ΕΠΛ646: Ενότητα Α

(Έλεγχος Ταυτοχρονίας)



```
Deposit (amount, account#) {  
    x = read(accounts[account#]);  
    write(accounts[account#], amount + x);  
}
```

Θεωρήστε:

Account[7] = €100

T1: Deposit1(100, 7)

T2: Deposit2(50, 7)

Εκτέλεση 1

Εκτέλεση 2

T1

T2

T1

T2

Read1
Write1

Read1

Read2

Write1

Read2
Write2

Write2

250€ (Correct)

150€ 😞!

WRONG

ΕΠΛ646: Ενότητα Α

(Τεχνικές Ανάκαμψης)



- Μια DBMS διασφαλίζει την **ατομικότητα** - ***atomicity*** (all-or-nothing) ακόμη και εάν το σύστημα καταρρεύσει στη μέση μιας δοσοληψίας.
- **Ιδέα:** Να διατηρείται ένα ***log*** (history) από όλες τις πράξεις που εκτελεί η DBMS καθώς εκτελεί ένα σύνολο δοσοληψιών:
 - **Προτού** οποιαδήποτε αλλαγή γίνει στην DB, το αντίστοιχο **log entry** εγγράφεται σε ασφαλές σημείο. (***WAL protocol***)
 - Μετά την κατάρρευση, οι επιδράσεις των ατελείωτων δοσοληψιών ακυρώνονται (***undone***) με τη χρήση του log (εάν δεν αποθηκεύτηκε το log entry τότε η αλλαγή δεν εφαρμόστηκε στη DB!)

ΕΠΛ646: Ενότητα Α

(Minibase)



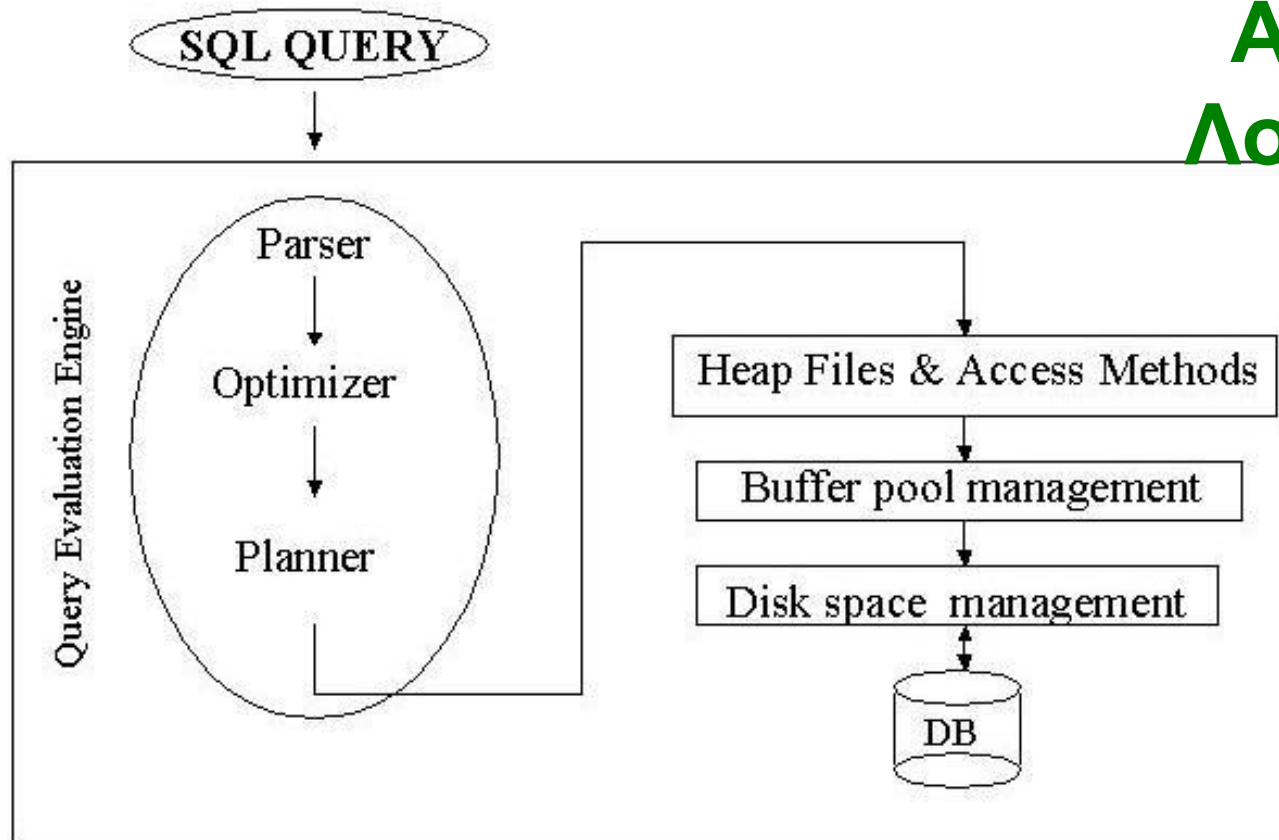
- Η **Minibase** είναι ένα σύστημα διαχείρισης δεδομένων το οποίο προορίζεται για εκπαιδευτική χρήση.
- Περιλαμβάνει ένα **Αναλυτή Επερωτήσεων (Parser)**, ένα **Βελτιστοποιητή Επερωτήσεων (Query Optimizer)**, **Διαχειριστή Ενδιάμεσης Μνήμης (Buffer Pool Manager)**, **Μηχανισμούς Αποθήκευσης** (heap files, secondary indexes based on B+ Trees), και **Διαχειριστή Μαγνητικού Δίσκου (Disk Space Manager)**.
- Επιτρέπει στο φοιτητή να προγραμματίσει συστατικά μιας βάσης με χρήση της C++.
- **Αναπτύχθηκε παράλληλα με ένα από τα βιβλία του μαθήματος μας!**
- Χρησιμοποιείται σαν εισαγωγικό εργαλείο εκπαίδευσης του προσωπικού από εταιρείες κατασκευής βάσεων δεδομένων (π.χ., oracle).

ΕΠΛ646: Ενότητα Α (Minibase Architecture)



MiniBase Structure

**Εύκολο &
Ανοικτό
Λογισμικό**

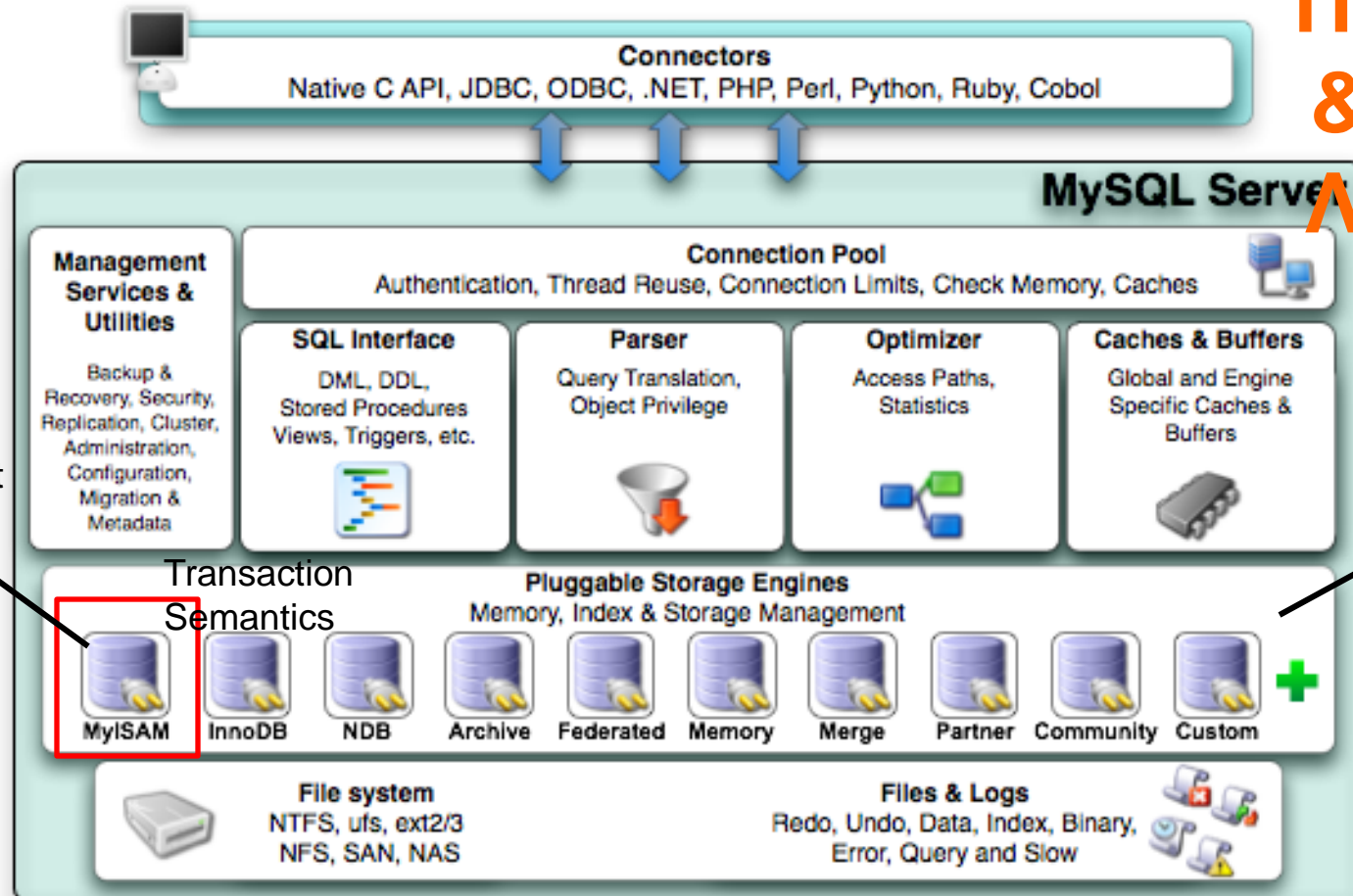


ΕΠΛ646: Ενότητα Α

(MySQL Server Architecture

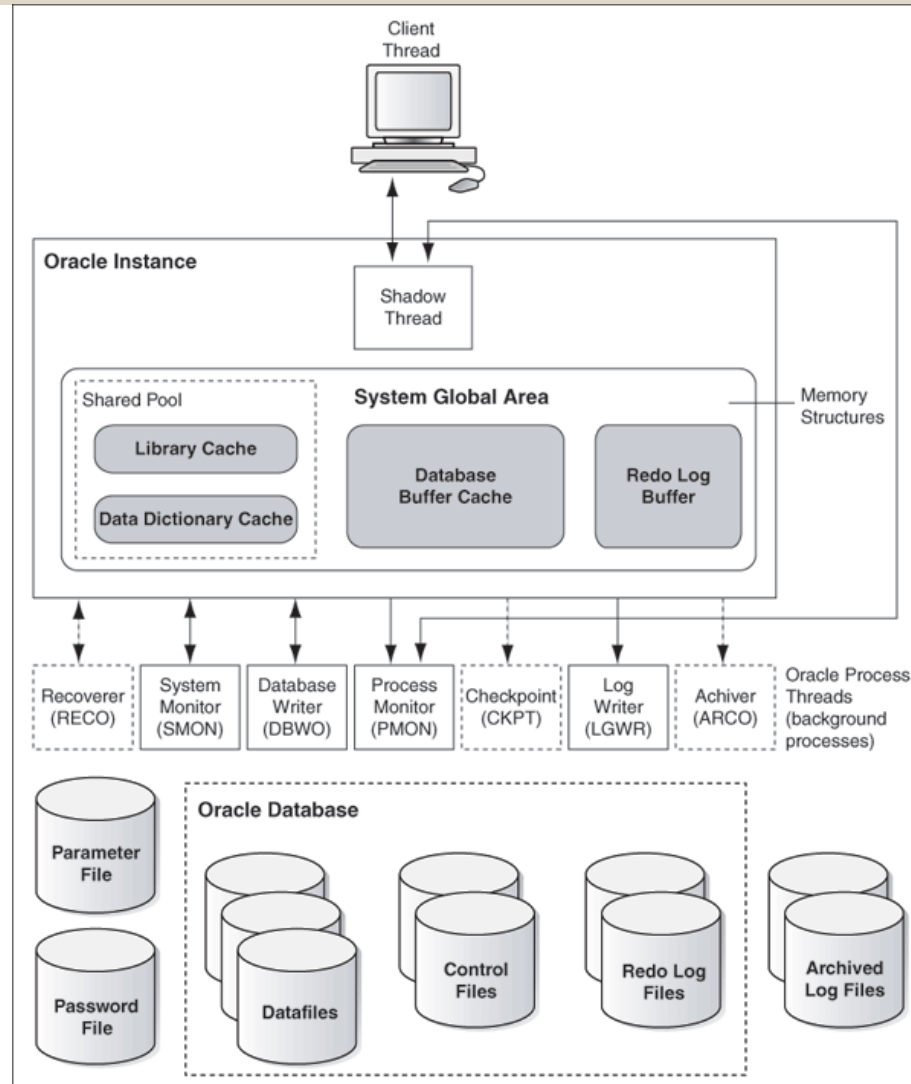


Περίπλοκο
& Ανοικτό
Λογισμικό



ΕΠΛ646: Ενότητα Α

(The Oracle Architecture)



**Περίπλοκο
& Κλειστό
Λογισμικό**

ΕΠΛ646: Ενότητα Β

Distributed/Web/Cloud DBs/Dstores



EPL342 – DBs (Modeling, SQL, Normalization)

Programmers / Users

SQL

EPL646 - PART A
(RDBMS Internals)

EPL646 - PART B
(Distributed/Web/Cloud DBs)



EPL646 - PART C
(Other DB Research)

ΕΠΛ646: Ενότητα Β

Distributed/Web/Cloud DBs/Dstores



- **Distributed Database (DDB)**

- a collection of multiple **logically related** (λογικά συσχετιζόμενες) **databases** distributed over a **computer network**.

- **Distributed Database Management System (DDBMS)**

- a **generic software system** that manages a **distributed database** while making the **distribution transparent** (διαφανής) to the user.

- **Applications:**

- Operational Scalability: OLTP Workloads
- Analytics (Business Intel.): OLAP Workloads

- All major vendors offer DDBMS extensions but there was never a common standard bringing vendors together.

ΕΠΛ646: Ενότητα Β

Distributed/Web/Cloud DBs/Dstores



"Big Data"

- *"Collection of **data sets so large and complex** that it becomes **awkward to work** with using on-hand **database management tools**." (wikipedia.org)*
- **Examples**
 - **Facebook** handles over 40 billion photos with **HBase**
 - **Google's Bigtable** is designed to scale into the petabyte range across "hundreds or thousands of machines, ...easy to add more machines ... without any reconfiguration".
 - **CERNs Large Hadron Collider (LHC)** produced 13 petabytes of data in 2010
 - **Walmart** handles more than 1 million customer transactions every hour (more than 2.5 petabytes of data = 167 times the info contained in all the books in the US Library of Congress.)

ΕΠΛ646: Ενότητα Β

Distributed/Web/Cloud DBs/Dstores



Google's Datacenter in Oregon



Microsoft's 224,000 Servers Only Take Four People To Set Up

- There are 2000 in that container. And there are 112 such containers in Microsoft's \$US500 million Chicago data centre

(<http://www.gizmodo.com.au/2009/10/microsofts-224000-servers-only-take-four-people-to-set-up/>)



ΕΠΛ646: Ενότητα Β

Distributed/Web/Cloud DBs/Dstores



- Γιατί οι RDBMS ΔΕΝ είναι κατάλληλες για **Big-data**;
 - **Ψηλό Κόστος**
 - Oracle Standard Edition (per CPU): **5,900\$**
 - Oracle Enterprise Edition (per CPU): **47,500\$**
 - IBM DB2 v9.7 Enterprise: **25,000\$**
 - SQL Server 2008 Enterprise: **25,000\$**
 - Τα πιο πάνω ΔΕΝ περιλαμβάνουν κόστος αγοράς υλικού (server), λειτουργικού συστήματος, training, κτλ.!
 - **Ψηλή Πολυπλοκότητα**
 - Οι Σχεσιακές ΒΔ έχουν περίπλοκη εσωτερική δομή (triggers, transactions, indexes, views, κτλ.) που δεν είναι χρήσιμα για τις εφαρμογές στα νέα αυτά περιβάλλοντα.
 - **Δεν παρέχουν Επεκτασιμότητα / Ελαστικότητα;**
 - Pay as you go?

ΕΠΛ646: Ενότητα Β

Distributed/Web/Cloud DBs/Dstores



NewSQL-as-a-Service

To Amazon RDS* (Relational Database Service)

Pay by the hour your DB Instance runs.

US – N. Virginia	US – N. California	EU – Ireland	APAC – Singapore
DB Instance Class			Price Per Hour
Small DB Instance			\$0.11
Large DB Instance			\$0.44
Extra Large DB Instance			\$0.88
Double Extra Large DB Instance			\$1.55
Quadruple Extra Large DB Instance			\$3.10

963\$ / year



27,165 \$ / year

DB Instance Classes

(*essentially MySQL running on Amazon EC2 – Elastic Computing Cloud)

Amazon RDS currently supports five DB Instance Classes:

- Small DB Instance: 1.7 GB memory, 1 ECU (1 virtual core with 1 ECU), 64-bit platform, Moderate I/O Capacity
- Large DB Instance: 7.5 GB memory, 4 ECUs (2 virtual cores with 2 ECUs each), 64-bit platform, High I/O Capacity
- Extra Large DB Instance: 15 GB of memory, 8 ECUs (4 virtual cores with 2 ECUs each), 64-bit platform, High I/O Capacity
- Double Extra Large DB Instance: 34 GB of memory, 13 ECUs (4 virtual cores with 3.25 ECUs each), 64-bit platform, High I/O Capacity
- Quadruple Extra Large DB Instance: 68 GB of memory, 26 ECUs (8 virtual cores with 3.25 ECUs each), 64-bit platform, High I/O Capacity

For each DB Instance class, RDS provides you with the ability to select from 5GB to 1TB of associated storage capacity. One ECU provides the equivalent CPU capacity of a 1.0-1.2 GHz 2007 Opteron or 2007 Xeon processor.

ΕΠΛ646: Ενότητα Β

Distributed/Web/Cloud DBs/Dstores



Το Amazon EC2 Σύστημα Διαπροσωπείας

The screenshot displays the AWS Management Console interface for the Amazon EC2 service. The 'My Instances' section is active, showing a list of instances. One instance, 'i-2392ff4a', is selected, and a context menu is open over it. The instance details show it is a Windows instance in the us-east-1a zone, with a status of 'running'.

Instance	AMI ID	Zone	Security Group	Type	Status	Platform	Key Pair
i-2392ff4a	ami-0529ce6c	us-east-1a	basicwin32	m1.small	running	windows	basicwin

1 EC2 Instance selected

Instance:	i-2392ff4a	Zone:	us-east-1a
AMI ID:	ami-0529ce6c	Type:	m1.small
Security Groups:	basicwin32	Owner ID:	arn:aws:iam::123456789012:user:ec2-user
Status:	running	Ramdisk ID:	-
Reservation:	r-8bf871e2		

© 2008, Amazon Web Services LLC or its affiliates. All right reserved. | Feedback | Support | Privacy Policy | Terms of Use | An amazon.com company

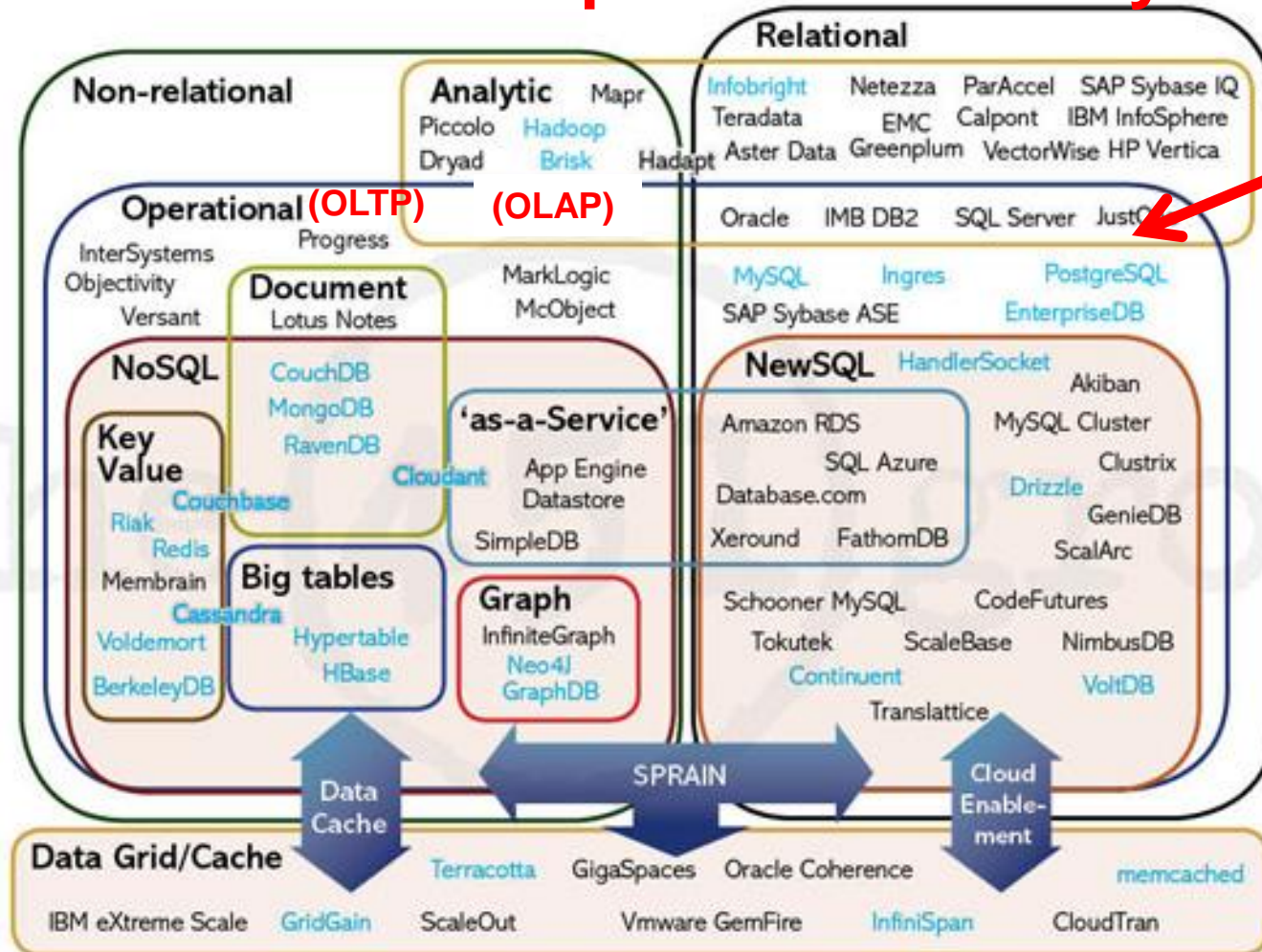
Στα πλαίσια του μαθήματος
θα χρησιμοποιήσουμε το
DMSL Vcenter ή VPS
(<http://dmsl.cs.ucy.ac.cy/>)

ΕΠΛ646: Ενότητα Β



Distributed/Web/Cloud DBs/Dstores

What is the picture like today?



<http://xeround.com/blog/2011/04/newsql-cloud-database-as-a-service>

ΕΠΛ646: Ενότητα Β



Distributed/Web/Cloud DBs/Dstores NoSQL

- A broad class of DBMSs that **Don't follow** the **relational** model (i.e., not using tables), thus those DBMSs are usually also not using SQL either.
- **Characteristics**
 - NoSQL, Distributed, Fault-tolerant Architectures, Less Consistency Guarantees, High Performance and High Scalability!
- **Examples**
 - Store/Analyze Google Maps (Bigtable), friendship data from Facebook (Cassandra, HBase), accounting data at Akamai (HBase), Amazon S3 (DynamoDB)

Big Data!

ΕΠΛ646: Ενότητα Β



Distributed/Web/Cloud DBs/Dstores

NewSQL



- **OLTP (Online Transaction Processing)**: facilitate & manage transaction-oriented applications (order something, withdraw money, cash a check, etc.)
- **New OLTP**: Consider new Web-based applications such as **multi-player games**, **social networking sites**, and **online gambling networks**.
 - The aggregate number of interactions per second is skyrocketing!.
- **New SQL**: An alternative to NoSQL or Old SQL for New OLTP applications.
- **Examples**: Clustrix, NimbusDB, and VoltDB.

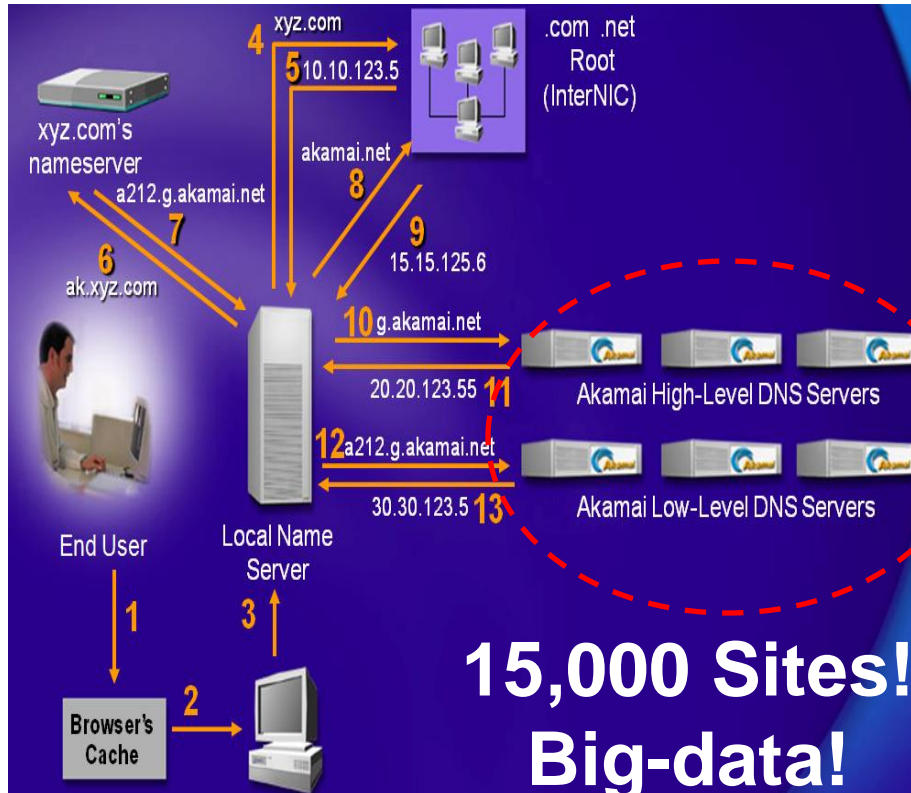
Big Data!

ΕΠΛ646: Ενότητα Β

Distributed/Web/Cloud DBs/Dstores



Big-data Example: Akamai Content Distribution Network



Big Data Akamai Jobs - Boston, MA | Careers Center - More information

<https://jobs-akamai.icims.com/jobs/6493/senior-system-software-engineer/job>

Location:	US-MA-Cambridge	Posted Date:	8/23/2012
Cost Center:	215	Category:	Engineering
ID:	6493		

Apply for this job:

Your application choices are:

- Apply for this job [online](#)

If you would like to include a cover letter, please be sure combine your cover letter and resume into one document.

More information about this job:

Overview:

About the Job
Be a Big Data software engineer. Extract statistics to provide insight into usage logs collected by Akamai's edge services. You will use our extensible distributed processing cluster to parse, aggregate and generate reports from logfiles with volume of more than Peta-Byte per day. You will work with the QA team to ensure that the resulting data products are highly accurate and available to all consumers.

ΕΠΛ646: Ενότητα Γ

Sensor/Spatio-temporal/etc.



EPL342 – DBs (Modeling, SQL, Normalization)

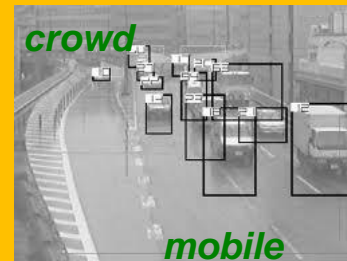
Programmers / Users

SQL

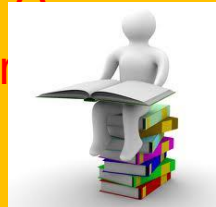
EPL646 - PART B
(Distributed/Web/Cloud DBs)

EPL646 - PART A
(RDBMS Internals)

EPL646 - PART C
(Other DB Research Topics)



Spatio-temporal



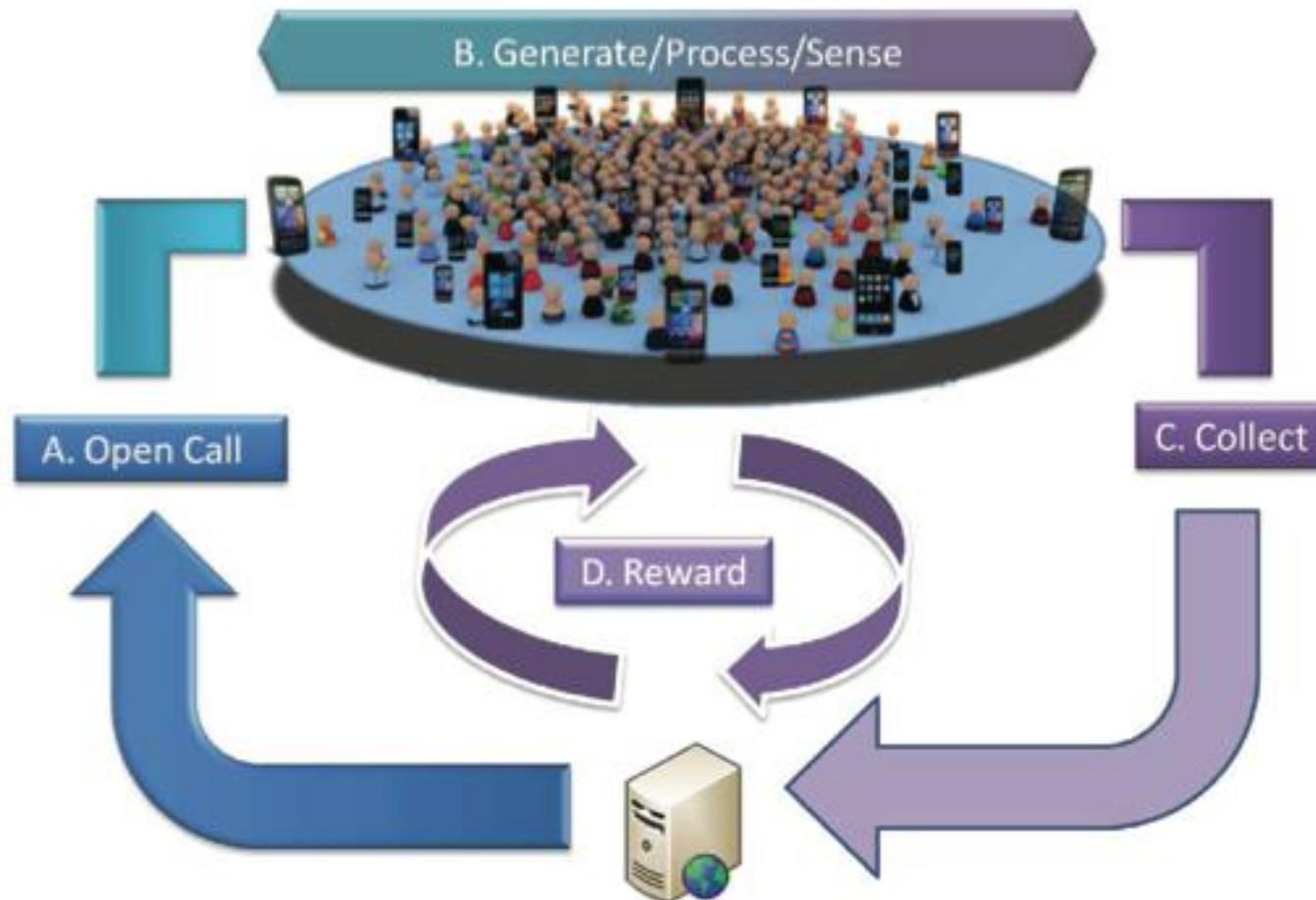
energy



sensor

ΕΠΛ646: Ενότητα Γ (Research)

Crowdsourced Databases

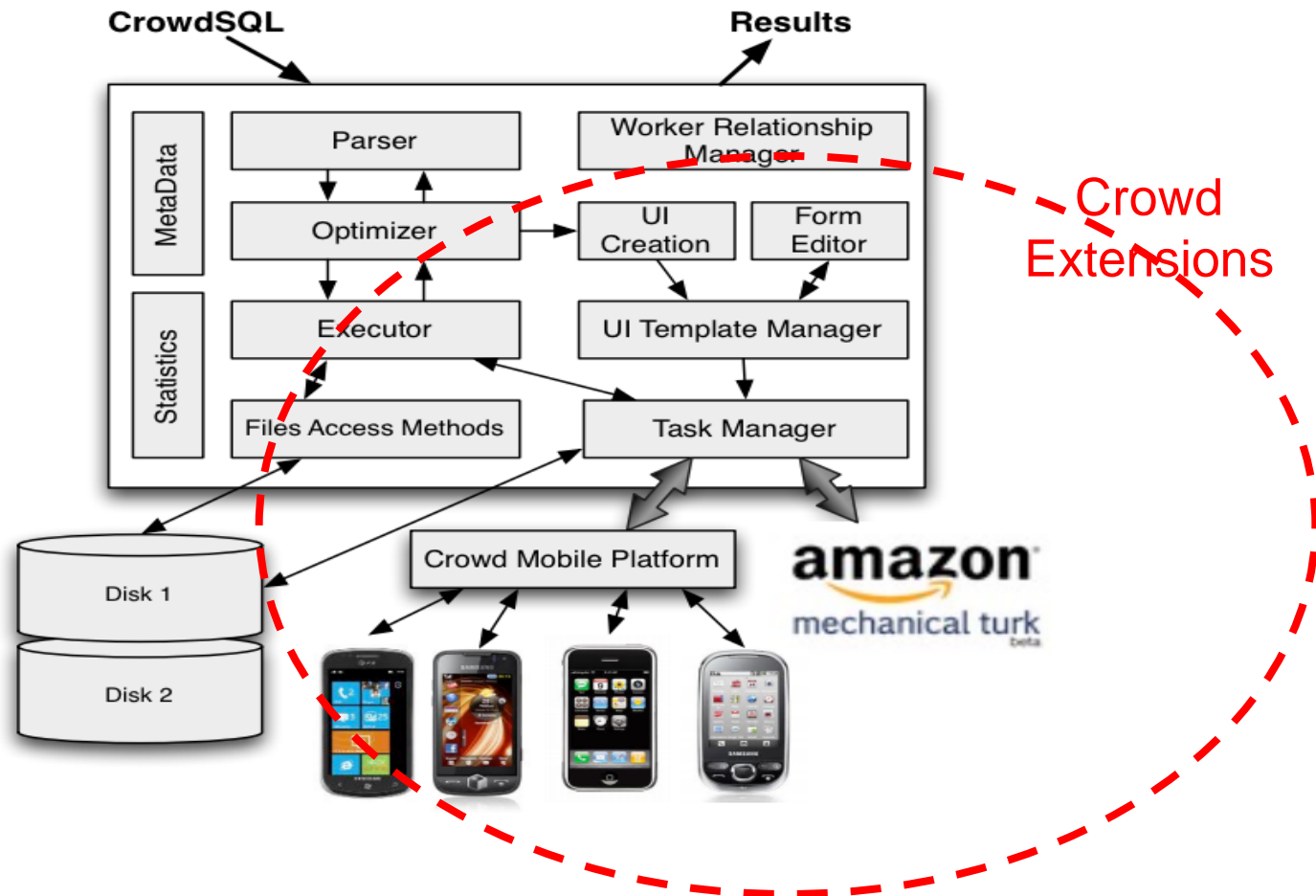


ΕΠΛ646: Ενότητα Γ (Research)

Crowdsourced Databases



- CrowdDB Architecture



ΕΠΛ646: Ενότητα Γ (Research)

Crowdsourced Databases



- **CrowdSQL in Action:**


- Finding Missing Data / Generate GUIs on-the-fly
 - `SELECT * FROM companies WHERE name = "IBM";`
- Fuzzy Matching (subjective comparison)
 - `SELECT * FROM companies WHERE`
 - `name ~="IBM";` (not regex-oriented LIKE)
- Fuzzy Ranking (subjective):
 - `SELECT image FROM pictures`
 - `ORDER BY novel_idea LIMIT 10`
- Fuzzy Aggregation (subjective), Join and typical DBMS operators (group-by, index scans, etc)
- CrowdDB: Answering Queries with Crowdsourcing, M. J. Franklin, D. Kossmann, T. Kraska, S. Ramesh, R. Xin, SIGMOD'11 & VLDB'11 Demo

Are the following entities the same?

IBM = Big Blue

(b) CROWDEQUAL

Which picture visualizes better "Golden Gate Bridge"

☐  ☐ 

ΕΠΛ646: Ενότητα Γ (Research)

Sensor Data Management



The screenshot displays the KSpot software interface, which is divided into three main panels:

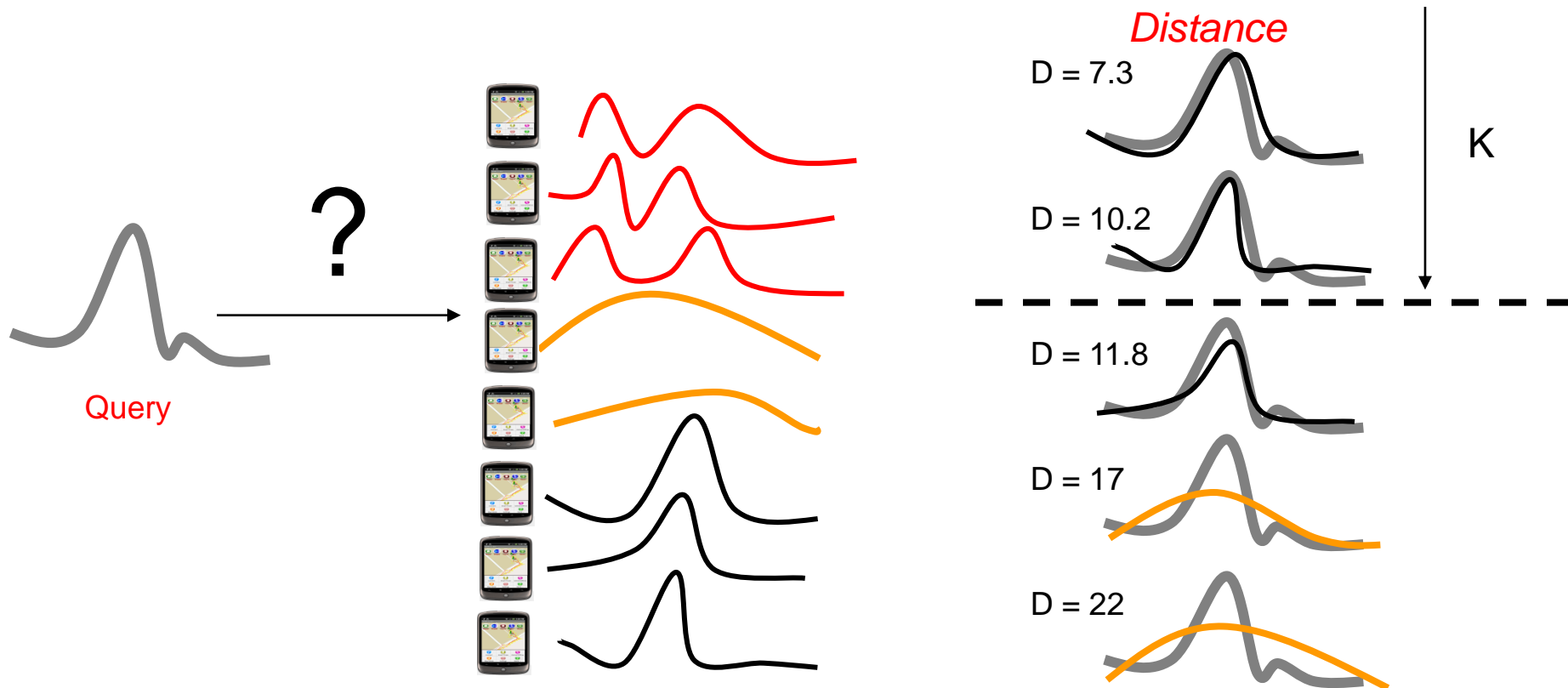
- KSpot Configuration Panel:** This panel is used for setting up the sensor network. It includes a 'Select Cluster' dropdown menu (currently set to 'Conf.Room B'), a 'Sensors Per Cluster' input field (set to 2), and a 'Sensors Configuration' table. A red arrow points to this panel with the label 'Configuration Panel'.
- KSpot Query Panel:** This panel is used for querying the sensor data. It includes a 'Select Query Type' dropdown (set to 'Current'), a 'Select TopK' input field (set to 3), and a table for selecting attributes and aggregates. A red dashed box highlights the 'Show Query Text' section, which contains the following SQL query:

```
SELECT TOP 3 cluster, AVG(sound)
FROM sensors
GROUP BY cluster
WITH HISTORY
```

 A red arrow points to this section with the label 'Query Box'.
- KSpot Display Panel:** This panel displays a floor plan of the building with sensor locations marked. Three red circles with numbers 1, 2, and 3 are overlaid on the floor plan, indicating specific sensor locations: 1 in the Lab, 2 in the Reception, and 3 in the Administration. A red arrow points from the 'Query Box' to the 'Reception' sensor location with the label 'Online Ranking'.

ΕΠΛ646: Ενότητα Β (Research)

Spatio-temporal Data Management





Student Presentations

...