Department of Computer Science
University of Cyprus

EPL646 – Advanced Topics in Databases

Lecture 1

Syllabus and Course Overview

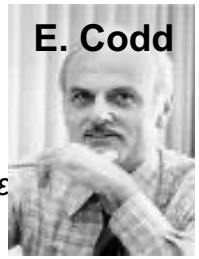
Demetris Zeinalipour
http://www.cs.ucy.ac.cy/~dzeina/courses/epl646

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).*
    - ***Χθές:** Το Σχεσιακό Μοντέλο υλοποιείται από τις περισσότερες σύγχρονες βάσεις δεδομένων αποτελώντας το υπόβαθρο των επιχειρήσεων)enterprise environments)*
    - ***Σήμερα:** Έντονη ανάγκη για μετάβαση σε νέες αρχιτεκτονικές οι οποίες υποστηρίζουν περισσότερες λειτουργίες (μηχανική μάθηση, ανάλυση δεδομένων & παρακολούθηση ροών) και προσφέρουν μεγαλύτερη Κλιμακωσιμότητα.*



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)

Today: 3rd largest after Microsoft, Alphabet

Server and Storage Systems

- Sun Servers
- Storage and Tape
- Exadata Database Machine
- SPARC SuperCluster T4-4
- Database Appliance
- Exalogic Elastic Cloud
- Oracle Solaris
- Oracle Linux
- Virtualization
- Enterprise Manager Ops Center
- More Servers and Storage ▾

Database

- Oracle Database 11*g*
- Real Application Clusters
- Data Warehousing
- Database Security
- Exadata Database Machine
- Database Appliance
- Big Data
- Enterprise Manager for Database
- Embedded
- High Availability
- MySQL
- More Database ▾

Middleware

- Java
- WebLogic Server
- Exalogic Elastic Cloud
- Exalytics In-Memory Machine
- SOA | BPM
- Social Network
- WebCenter
- Content | Portal
- Business Analytics
- Identity Management
- Enterprise Manager for Middleware
- Data Integration
- More Middleware ▾

Applications

- Oracle Fusion Applications
- Oracle E-Business Suite
- Human Capital Management
- PeopleSoft (HRM)
- RightNow
- Siebel (CRM)
- ATG
- Oracle CRM On Demand
- JD Edwards EnterpriseOne
- JD Edwards World
- Hyperion
- Primavera
- Application Integration
- More Applications ▾

Industries

- Communications
- Financial Services
- Healthcare
- High Technology
- Insurance
- Life Sciences
- Public Sector
- Retai
- Utilitie
- More

Τι καλύπτει το ΕΠΛ646;

EPL342 – DBs (Modeling, SQL, Normalization)

Programmers / Users

SQL

EPL646 - PART A
(RDBMS Internals)

DBMS

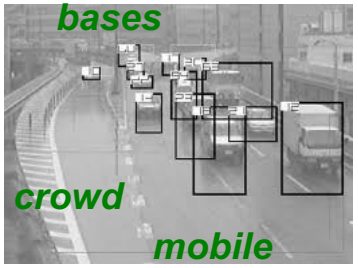
EPL646 - PART B
(Distributed/Web/Cloud DBs)

EPL646 - PART C
(Other DB Research)

Knowledge bases

streams

indoor

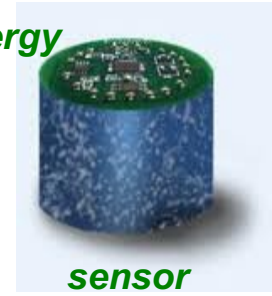


energy

crowd

mobile

Spatio-temporal



sensor

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 (Παρουσίαση / Συζήτηση Άρθρων - Νέα Ώρα)
  - Εργαστήριο (2 ώρες εβδομαδιαίως)
- **Υπόβαθρο**
  - Επαρκή γνώση σε συστήματα Linux (ΕΠΛ421) και προγραμματισμός σε γλώσσες C/C++/JAVA (ΕΠΛ232)

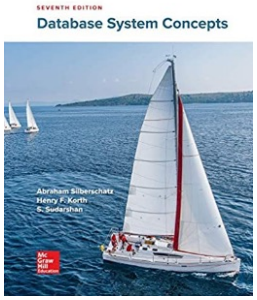
Συμβόλαιο Μαθήματος

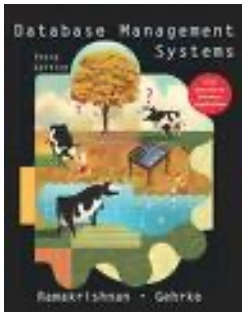
- **Αξιολόγηση**
  - 50% Τελική Εξέταση (1)
  - 20% Ενδιάμεση Εξέταση (1)
    - **Προκαταρτική Ημερομηνία:**

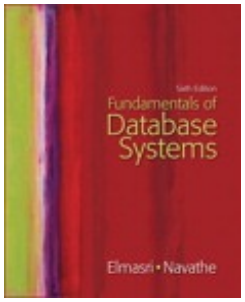
      Friday, 10/3/23 (8η βδομάδα)!
  - 30% Ασκήσεις
    - Προγραμματιστικές/Θεωρητικές Ασκήσεις
    - Παρουσιάσεις Άρθρων

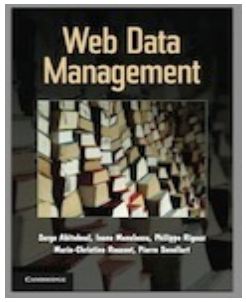
Βιβλιογραφία

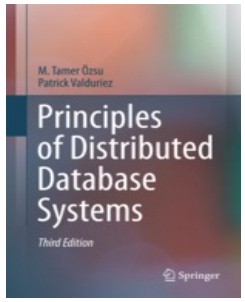
- Σημειώσεις Μαθήματος και Επιλεγμένη Αρθρογραφία
- **Database System Concepts,** 7th Edition, by Abraham Silberschatz, Henry Korth, S. Sudarshan, McGraw Hill; 7th edition, 1376 pages, ISBN-10 : 0078022150, **2019**.
- **Fundamentals of Database Systems**, 7/E Ramez Elmasri, Shamkant B. Navathe, ISBN-10: 0133970779, ISBN-13: 9780133970, **2016**
- **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.







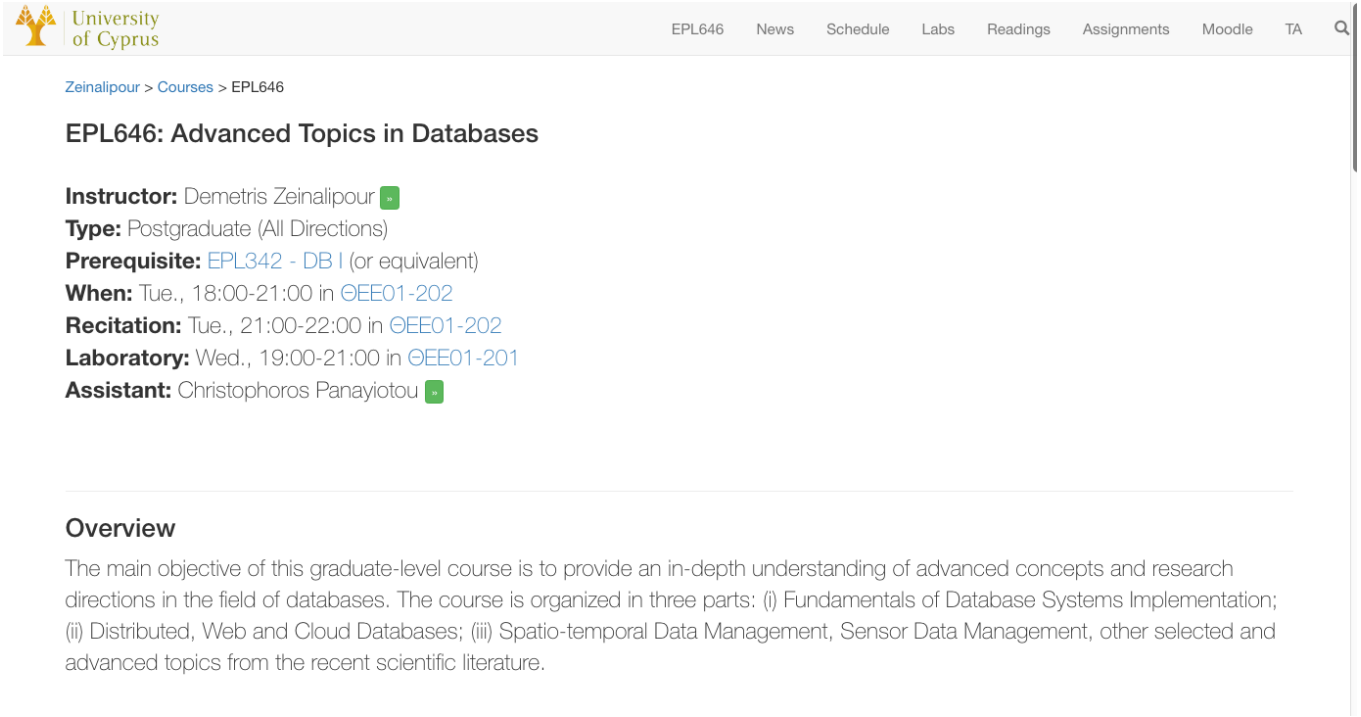




WWW

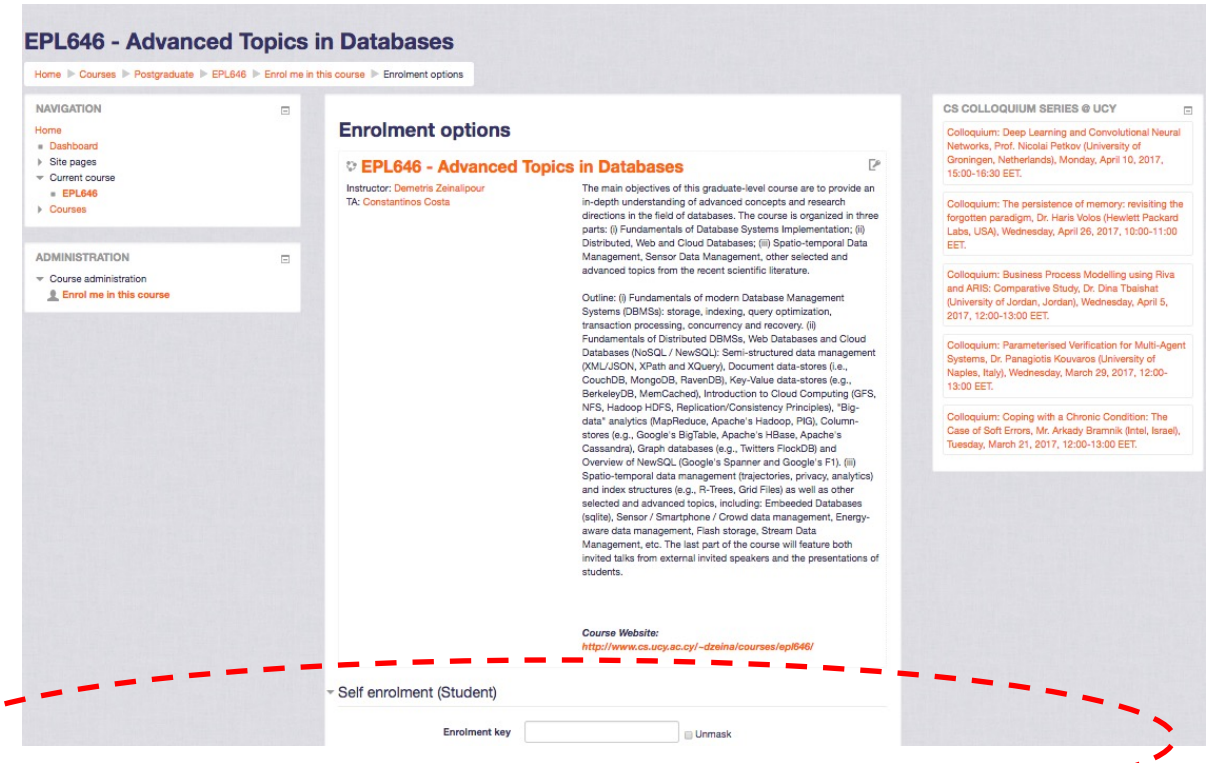
Πληροφορίες σχετικά με το μάθημα:

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



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/



ΕΠΛ646: Ενότητα Α
Εσωτερική Λειτουργία ενός RDBMS

EPL342 – DBs (Modeling, SQL, Normalization)

Programmers / Users

SQL

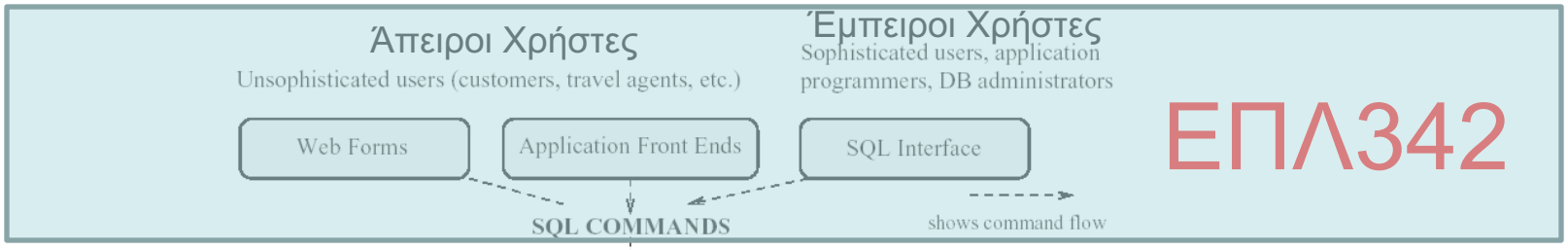
EPL646 - PART A
(RDBMS Internals)

DBMS

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

EPL646 - PART C
(Other DB Research)

ΕΠΛ646: Ενότητα Α
Εσωτερική Λειτουργία ενός RDBMS

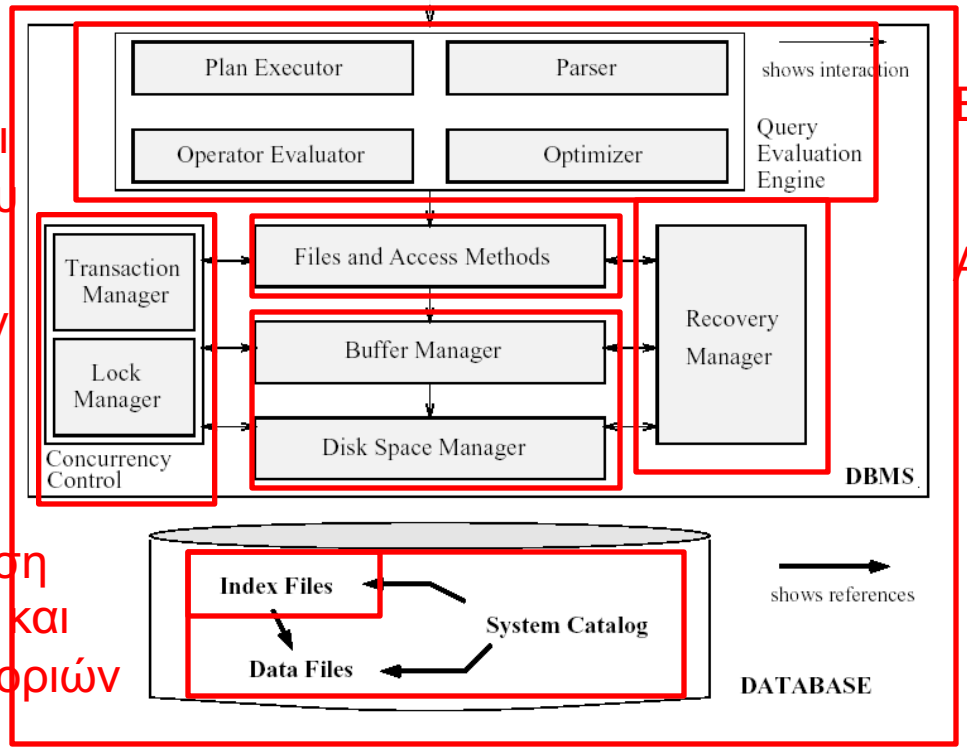


ΕΠΛ342

Έννοιες Δοσοληψιών και Τεχνικές Ελέγχου Ταυτοχρονίας

Δομή Ευρετηρίων Δευτερεύουσας Μνήμης (Hash, B+)

Αποθήκευση Δεδομένων και Μετα-πληροφοριών



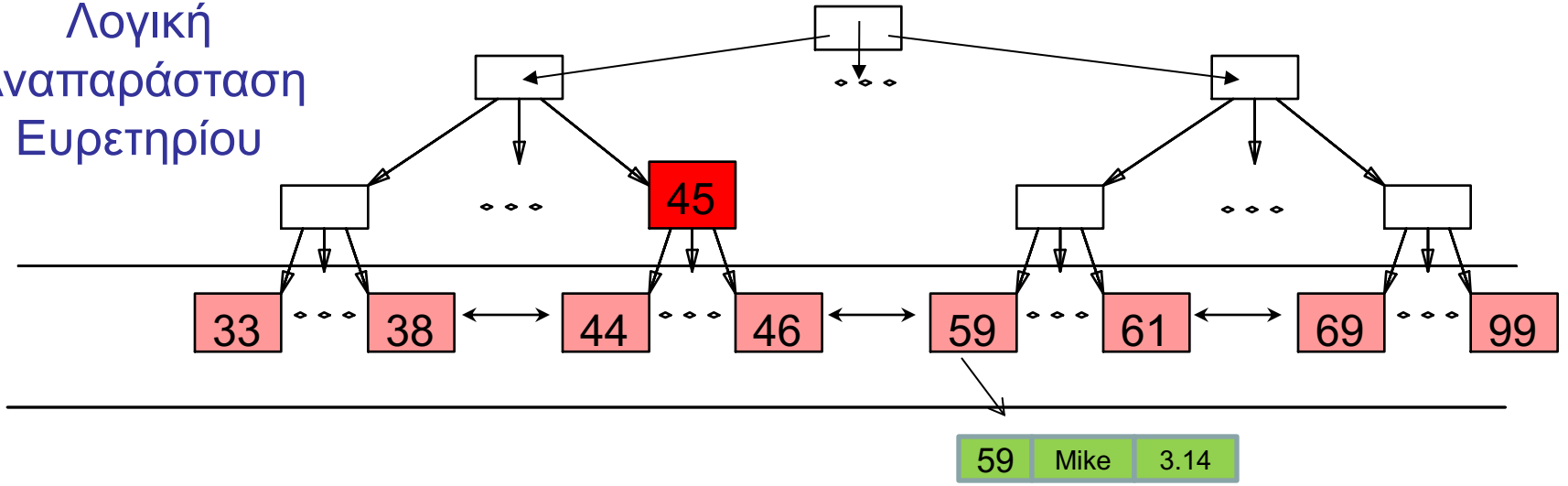
Αλγόριθμοι Βελτιστοποίησης Επερωτήσεων

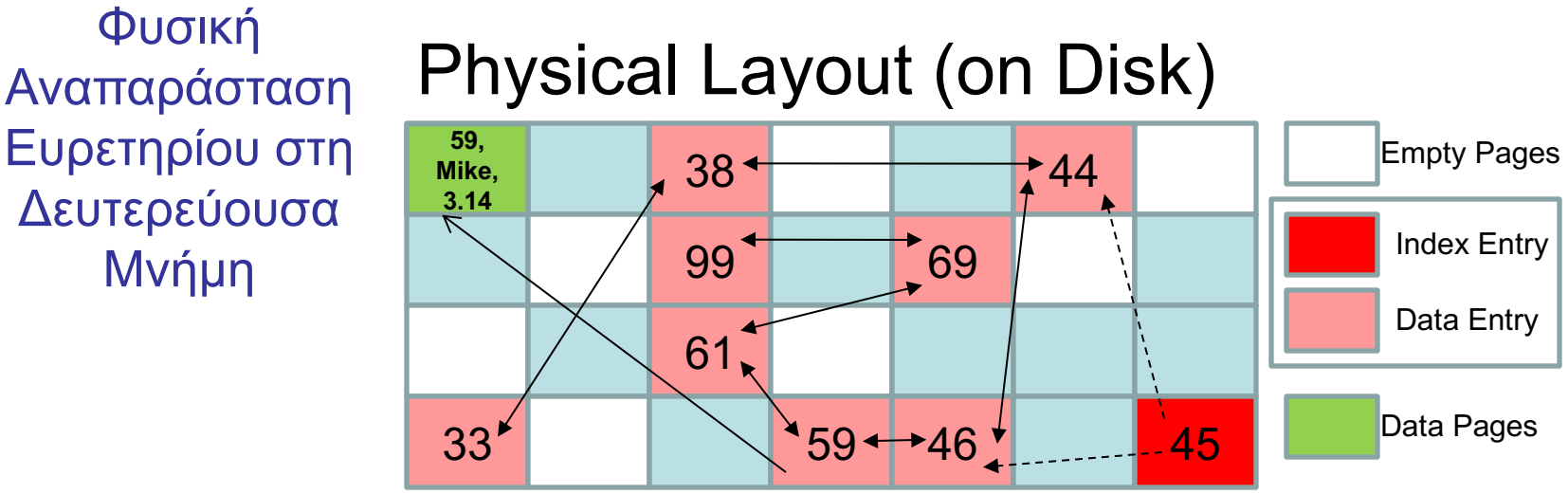
Τεχνικές Ανάκαμψης (σε περιπτώσεις σφαλμάτων)

Ενδόμνημη Διαχείριση Δεδομένων

ΕΠΛ646

ΕΠΛ646: Ενότητα Α
((Disk-based) Index Structures)





ΕΠΛ646: Ενότητα Α (Βελτιστοποίηση Επερωτήσεων)

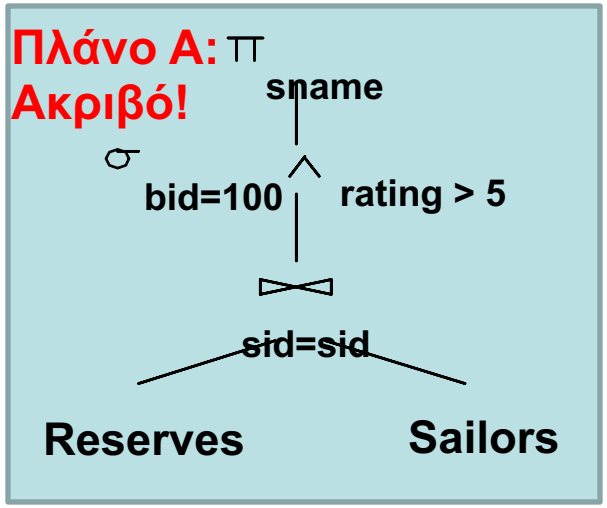
Αναλυτής (Parser): Αναλύει τα SQL επερωτήματα του χρήστη και τα μεταφέρει στον Βελτιστοποιητή

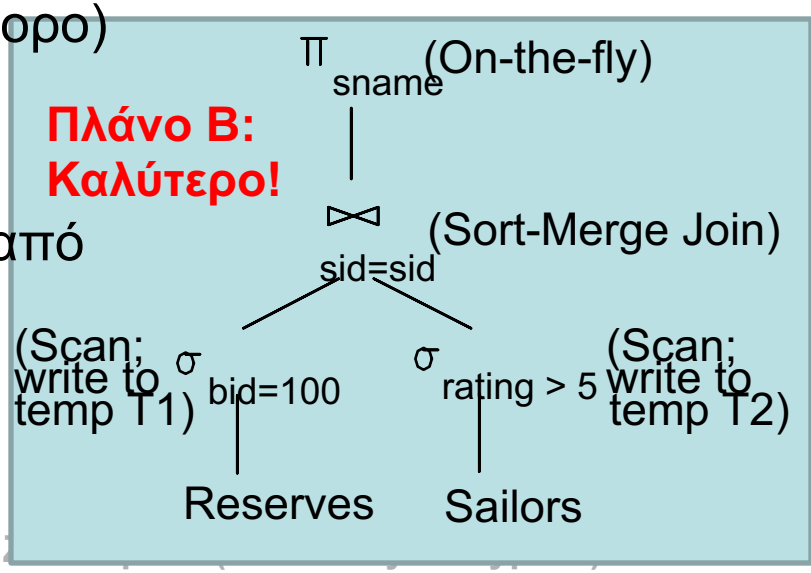
Βελτιστοποιητής (Optimizer): Κάνει χρήση μέτα-πληροφοριών στον **κατάλογο συστήματος (system catalog)** για να γνωρίζει τα διαθέσιμα ευρετήρια, τον αριθμό των πλειάδων σε ένα πίνακα.

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). Αυτό διασφαλίζεται από τα transactions.

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

T1	T2
Read1	
Write1	
	Read2
	Write2

250€ (Correct)

Εκτέλεση 2

T1	T2
Read1	
	Read2
Write1	
	Write2

150€ ☹!

WRONG

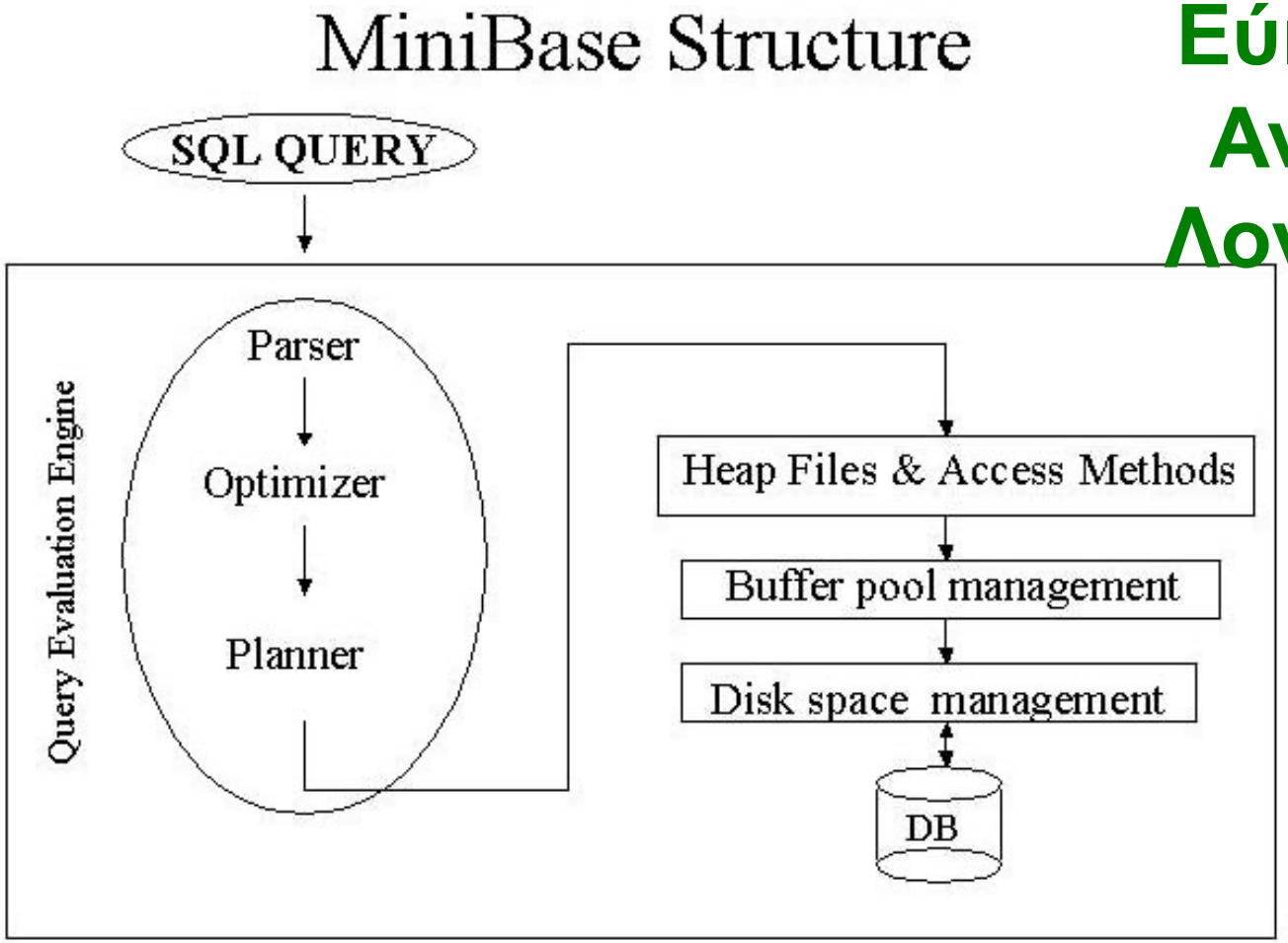
ΕΠΛ646: Ενότητα Α (Τεχνικές Ανάκαμψης)

- 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)

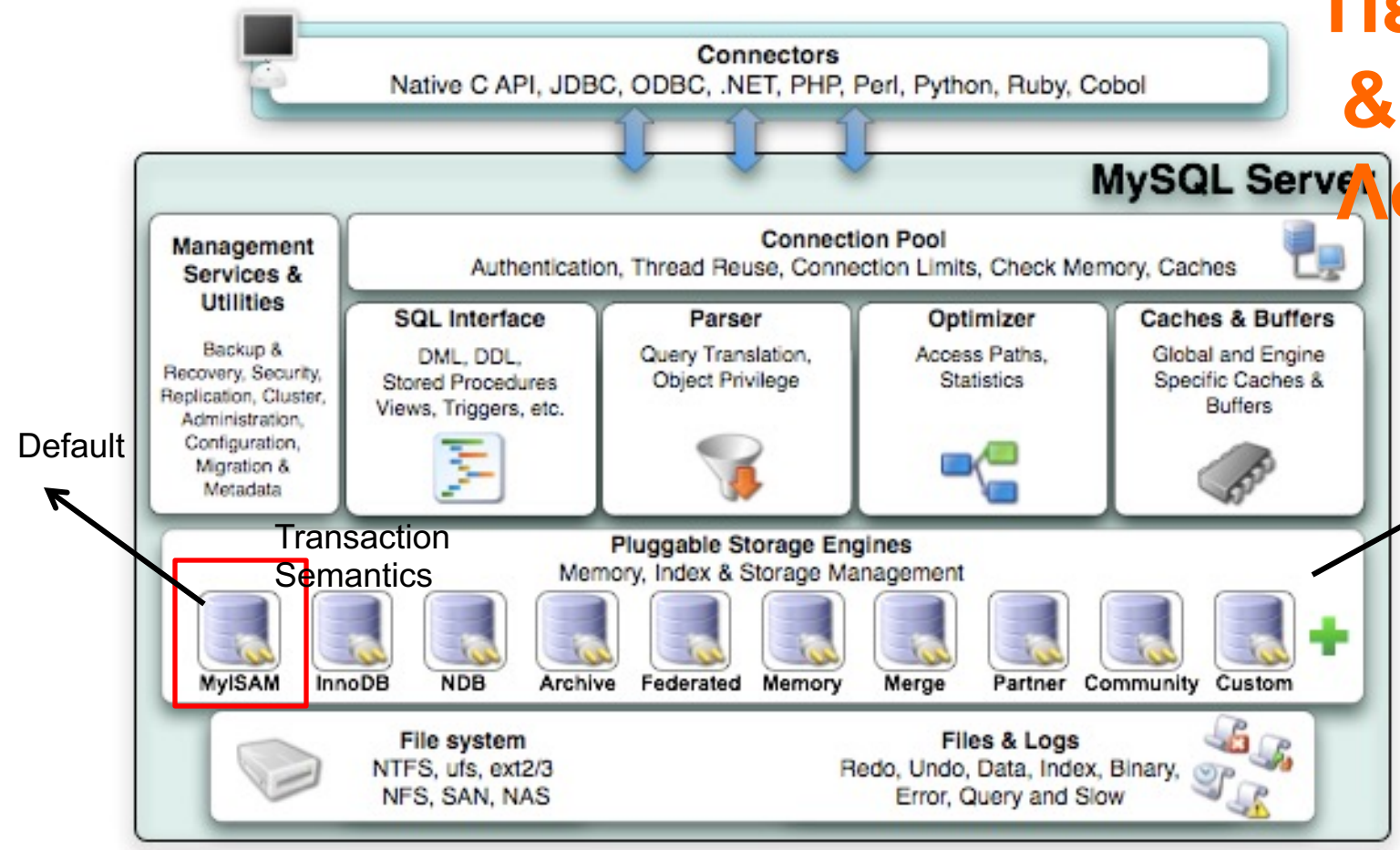
- Η **Minibase** είναι ένα σύστημα διαχείρισης δεδομένων το οποίο προορίζεται για εκπαιδευτική χρήση.
- Περιλαμβάνει ένα ***Αναλυτή Επερωτήσεων (Parser)***, *ένα* ***Βελτιστοποιητή Επερωτήσεων*** *(Query Optimizer),* ***Διαχειριστή Ενδιάμεσης Μνήμης*** *(Buffer Pool Manager),* ***Μηχανισμούς Αποθήκευσης*** *(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) πριν διεισδύσουν σε πιο περίπλοκο κώδικα (π.χ., postgres).*

ΕΠΛ646: Ενότητα Α (Minibase Architecture)



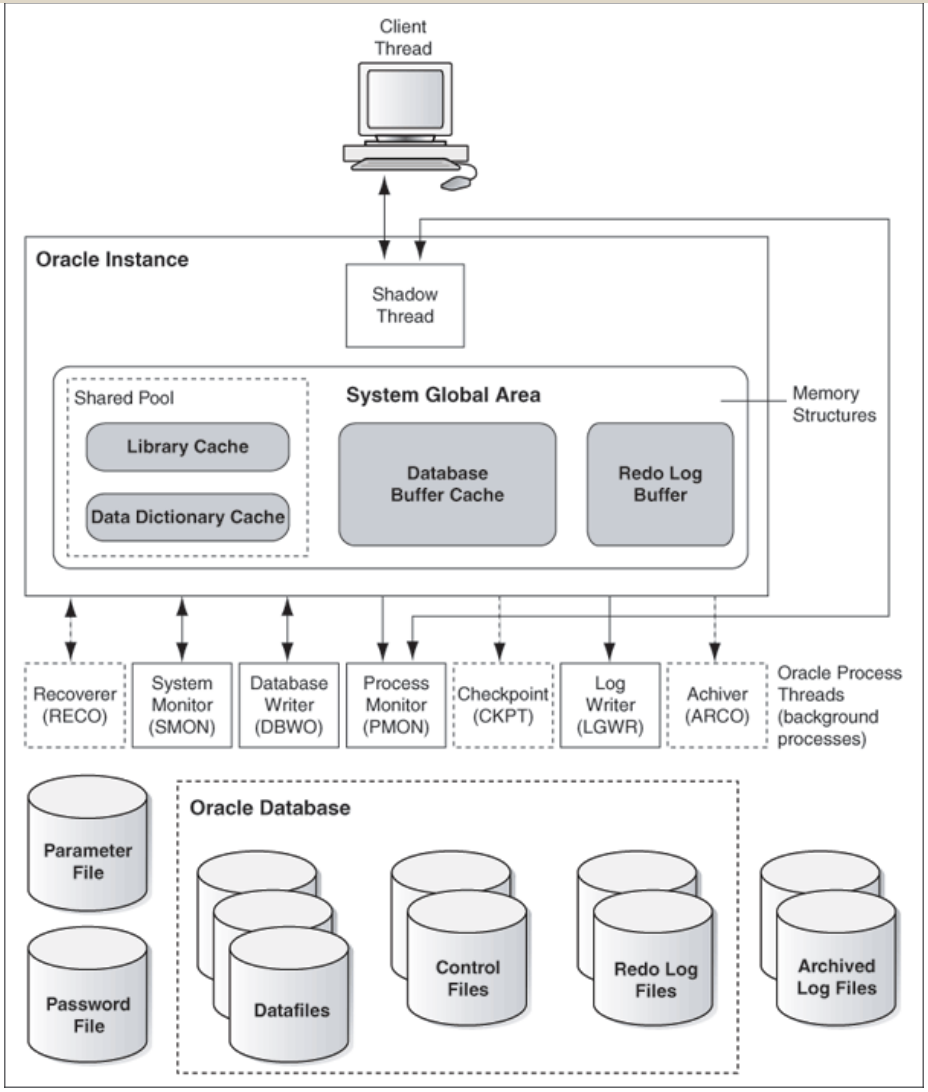
Εύκολο & Ανοικτό Λογισμικό

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 B
(Distributed/Web/Cloud DBs)

EPL646 - PART A
(RDBMS Internals)

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.

"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

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

DB Instance Classes

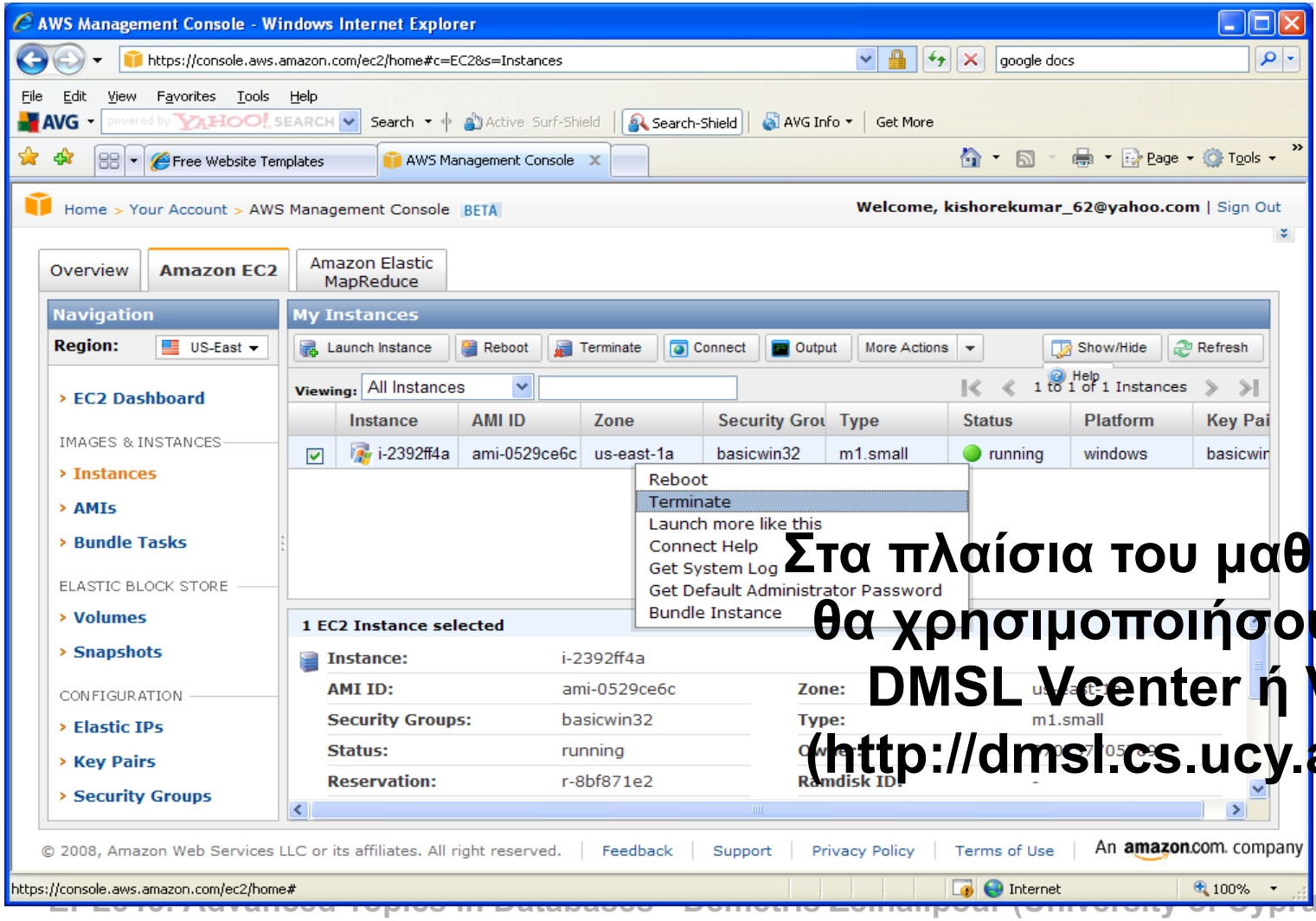
Amazon RDS currently supports five DB Instance Classes:

- 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

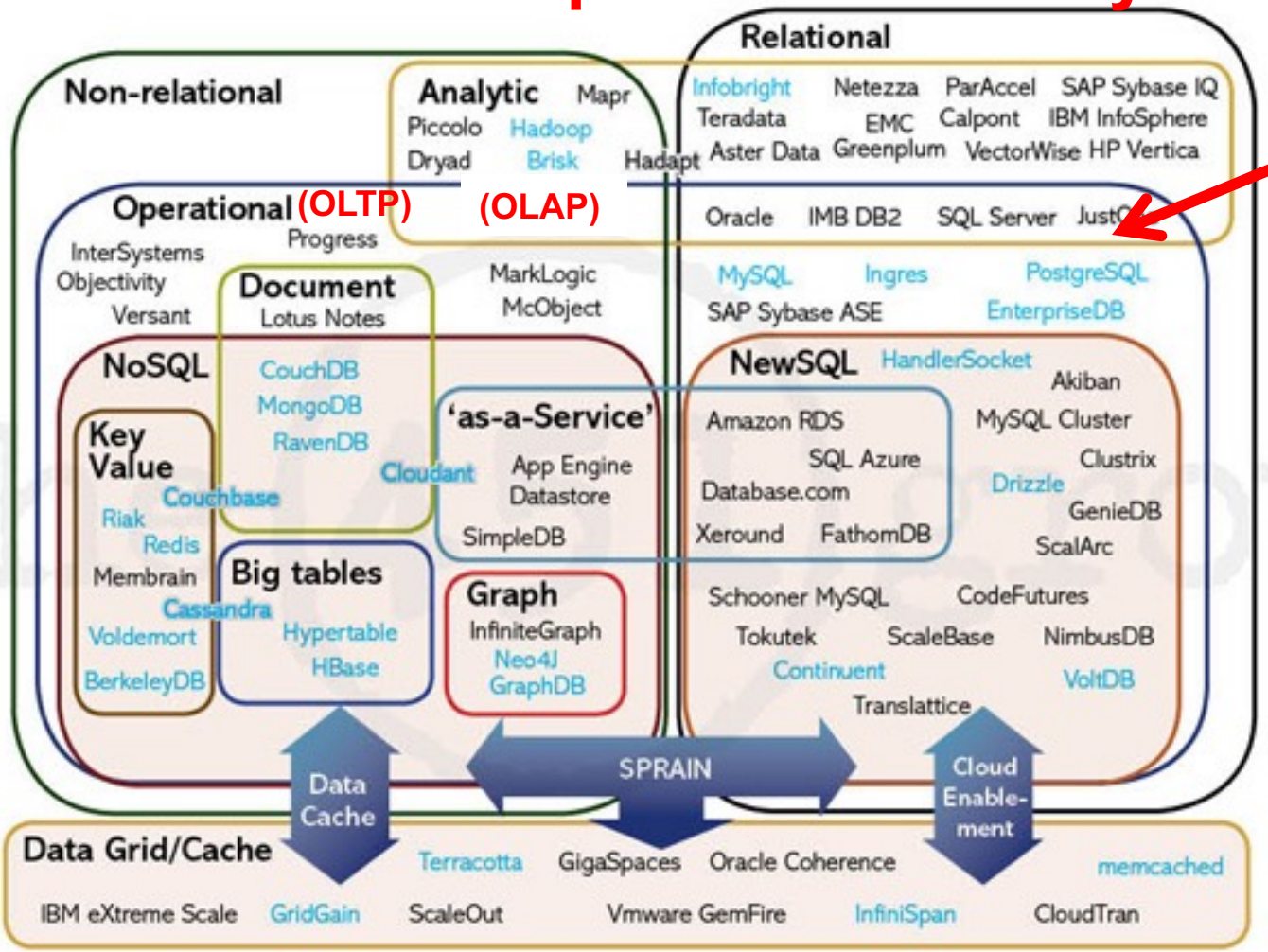
Το Amazon EC2 Σύστημα Διαπροσωπείας



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?



Traditional RBMSs

Venn Diagram by 451 group

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

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

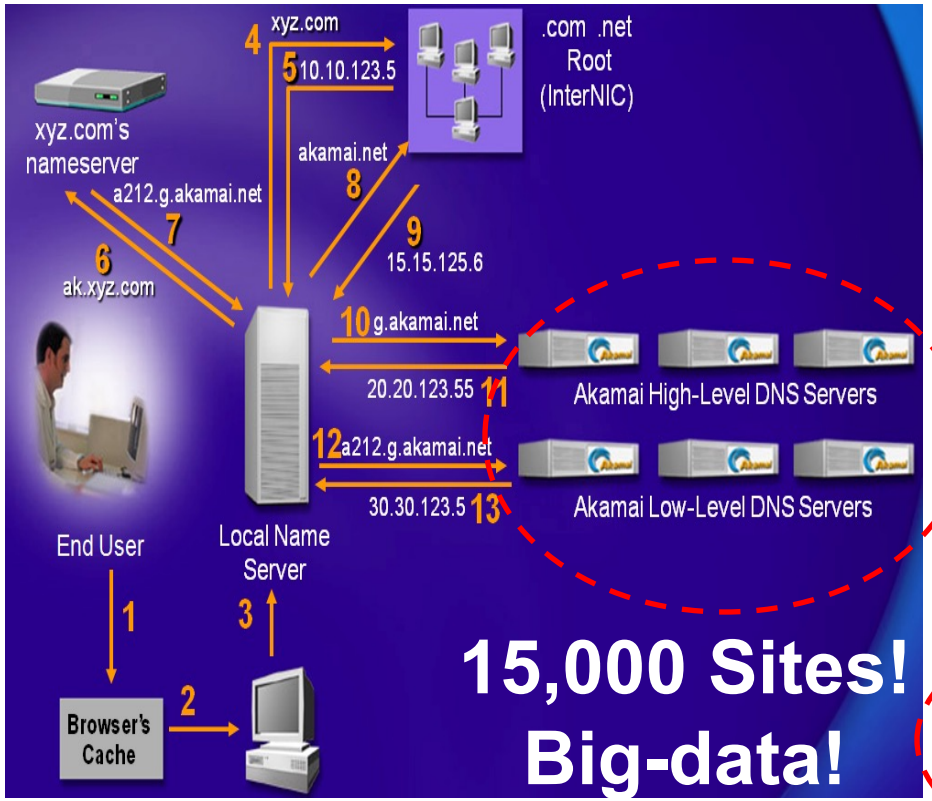
- **OLTP (Online Transaction Processing**): facilitate & manage transaction-oriented applications (order something, withdraw money, cash a check, etc.)
- **New OLTP:** Consider new Web-based applications such as **multi-player games**, **social networking** sites, and **online gambling networks**.
  - 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.

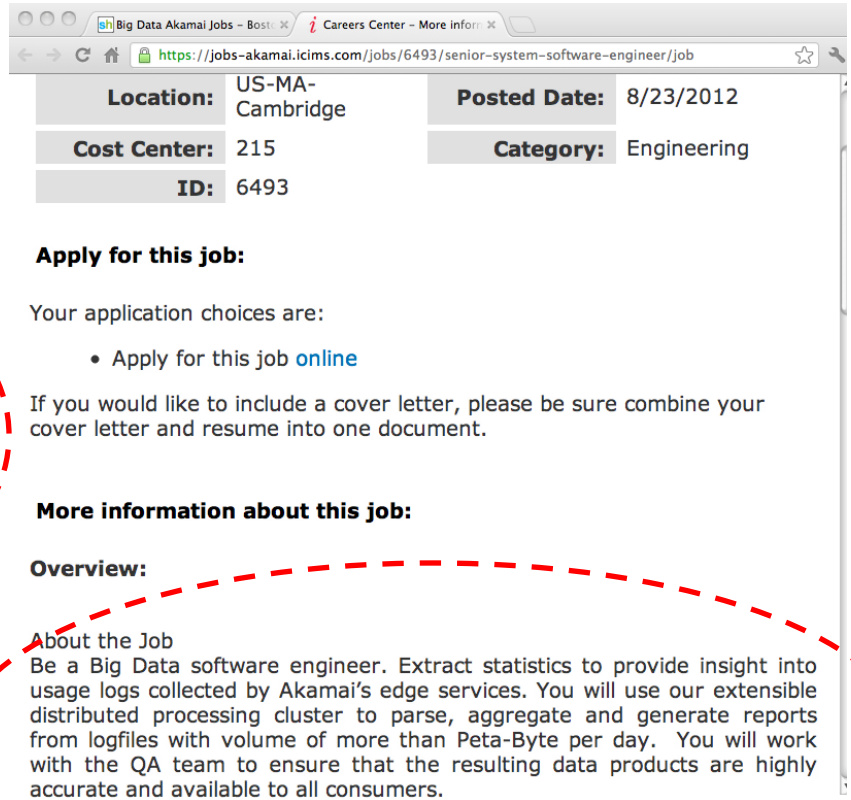
Big Data!

** Michael Stonebraker, June 16, 2011, http://tinyurl.com/9fok4kt*

ΕΠΛ646: Ενότητα Β
Distributed/Web/Cloud DBs/Dstores

Big-data Example: Akamai Content Distribution Network





Big Data Akamai Jobs – Bost | Careers Center – More infor

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

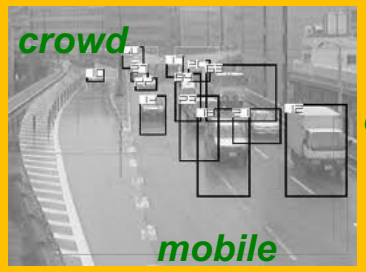
SQL

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

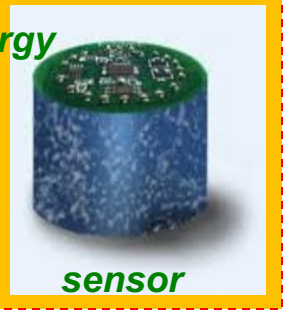
EPL646 - PART A
(RDBMS Internals)

EPL646 - PART C
(Other DB Resear

streams



energy



Spatio-temporal

VLDB - INTERNATIONAL CONFERENCE ON VERY LARGE DATA BASES

https://vldb.org/2023/

- **Data Privacy and Security**
  - Blockchain
  - Access control and privacy
- **Data Mining and Analytics**
  - Mining/analysis of different types of data (e.g., scientific/business, social networks, streams, text, web, graphs, rules, patterns, logs, and spatio-temporal)
  - Data warehousing, OLAP, parallel and distributed data mining
- **Distributed Database Systems**
  - Data networking and content delivery
  - Cloud data management, resource management, database as a service
  - Distributed transactions
  - Distributed analytics
- **Database Engines**
  - Currency control, recovery, and transactions
  - Access methods
  - Multi-core processing and hardware acceleration
  - Memory and storage management
  - Views, indexing, and search
  - Query processing and optimization
  - Administration and manageability
  - Database Performance and Manageability
  - Tuning, benchmarking, and performance measurement

- **Information Integration and Data Quality**
  - Heterogeneous and federated DBMS, metadata management
  - Data cleaning, data preparation
  - Schema matching, data integration
  - Knowledge graphs and knowledge management
  - Web data management and Semantic Web
  - Source discovery
- **Graph and Network Data**
  - Hierarchical, non-relational, and other modern data models
  - Graph data management
  - Social networks
- **Machine Learning, AI, and Databases**
  - Data management issues and support for ML and AI
  - Applied ML and AI for data management
- **Languages**
  - Schema management and design
  - Data models and query languages
- **Provenance and Workflows**
  - Process mining
  - Debugging
  - Provenance analytics
  - Profile-based and context-aware data management
- **Novel Database Architectures**
  - Embedded and mobile databases
  - Data management on novel hardware
  - Real-time databases, sensors and IoT, stream databases
  - Energy-efficient data systems
  - Video management and analytics systems
- **Text and Semi-Structured Data**
  - Information retrieval
  - Data extraction
  - Text in databases
  - Semi-structured data management, RDF

- **Specialized and Domain-Specific Data Management**
  - Ethical data management
  - Crowdsourcing
  - Image and multimedia databases
  - Fuzzy, probabilistic, and approximate data
  - Spatial and temporal databases
  - Scientific and medical data management
- **User Interfaces**
  - Database support for visual analytics
  - Data exploration tools
  - Database usability

IEEE ICDE - International Conference on Data Engineering

- AI for Database Systems
- Benchmarking, Performance Modeling, Tuning, and Testing
- Cloud Data Management
- Crowdsourcing
- Data Mining and Knowledge Discovery
- Data Models, Semantics, Query languages
- Data Stream Systems and Edge Computing
- Data Visualization and Interactive Data Exploration
- Database Security and Privacy
- Database technology for AI
- Database technology for Blockchains
- Distributed, Parallel and P2P Data Management
- Explainability, Fairness, and Trust in Data Systems and Analysis
- Graphs, Networks, and Semistructured Data
- Information Integration and Data Quality
- IoT Data Management
- Modern Hardware and In-Memory Database Systems
- Query Processing, Indexing, and Optimization
- Spatial Databases and Temporal Databases
- Text, Semi-Structured Data, IR, Image, and Multimedia databases
- Uncertain, Probabilistic, and Approximate Databases
- Very Large Data Science Applications/pipelines
- Workflows, Scientific Data Management

https://icde2023.ics.uci.edu/

ACM SIGMOD - International Conference on Management of Data

https://2023.sigmod.org/

- Benchmarking, database monitoring, and performance tuning
- Cloud data management and HPC
- Crowdsourced and collaborative data management
- Data models and semantics
- Data provenance and workflows
- Data exploration, visualization, query languages, and user interfaces
- Data integration, information extraction, and schema matching
- Data quality, data cleaning, and database usability
- Data warehousing, OLAP, SQL Analytics
- Data security, privacy, and access control
- Data sparsity, boosting, simulated data, and digital twins
- Data platforms for emerging hardware/Emerging hardware for data management
- Data systems for knowledge discovery, data mining, machine learning, and artificial intelligence
- Distributed, decentralized, and parallel data management, distributed ledgers, and blockchainsGraphs, social networks, and semantic web
- Machine learning and artificial intelligence for data management and data systems
- Multimedia and information retrieval
- Query processing and optimization
- Responsible data management and data fairness
- Self-driving databases
- Semistructured, partially structured, and unstructured data
- Sensor networks and IoT
- Spatial data management
- Storage, indexing, and physical database design
- Streams and complex event processing
- Temporal databases
- Transaction processing
- Uncertain, probabilistic, and approximate databases

Hosts yearly the SIGMOD Programming Competition!

IEEE MDM - IEEE International Conference on Mobile Data Management

https://mdmconferences.org/mdm2023/

- Mobility Data Acquisition and Protection
- Middleware and Tools for Mobile and Pervasive Computing
- Quality of mobility data: methodologies, metrics, algorithms
- Mobility Simulation
- Security, Privacy and Ethics in Mobility Data and Analytics
- Mobility Data Management and Processing
- Theoretical Foundations of Data-intensive Mobile Computing
- Data Management for Internet of Things (IoT) and Sensor Systems
- Mobile Crowd-Sourcing and Crowd-Sensing
- Data Stream Processing in Mobile/Sensor Network
- Indexing, Optimization and Query Processing for Moving Objects/Users
- Mobile Systems
- Mobile Cloud Computing and Data Management in the Mobile Cloud
- Mobile Location-Based Social Networks
- Mobile Recommendation Systems
- Context-aware Computing for Intelligent Mobile Services
- Learning and analytics
- Approaches
- Mobile Data Analytics
- Machine Learning/AI for Mobile Data
- Visual Analytics
- Behavioral/Activity Sensing and Analytics
- Applications of Mobility Data Science
- Data Management for Connected Cars, Intelligent Transportation Systems, Smart Spaces
- Routing, Personalized Routing, Eco-Routing, Routing for Electrical Vehicles
- Transportation-As-A-Service, Mobility-As-A-Service
- Data Management for Augmented Reality Systems
- Innovative Applications driven by Mobile Data
- Connections of mobile data management with other emergi technologies such as blockchain and paradigms, such as social sciences.
- Data Economy, Incentive Mechanisms, Reputation Systems and Game-theoretic

IEEE COMPUTER SOCIETY

Student Presentations

. . .