



Εθνικό και Καποδιστριακό Πανεπιστήμιο Αθηνών  
Τμήμα Πληροφορικής & Τηλεπικοινωνιών

# Πληροφορικά Συστήματα

7ο Εξάμηνο 2020-21

Θέματα διαχείρισης δεδομένων

Δρ. Κώστας Σαΐδης ([saiko@di.uoa.gr](mailto:saiko@di.uoa.gr))

# Θέματα διαχείρισης δεδομένων στα ΠΣ

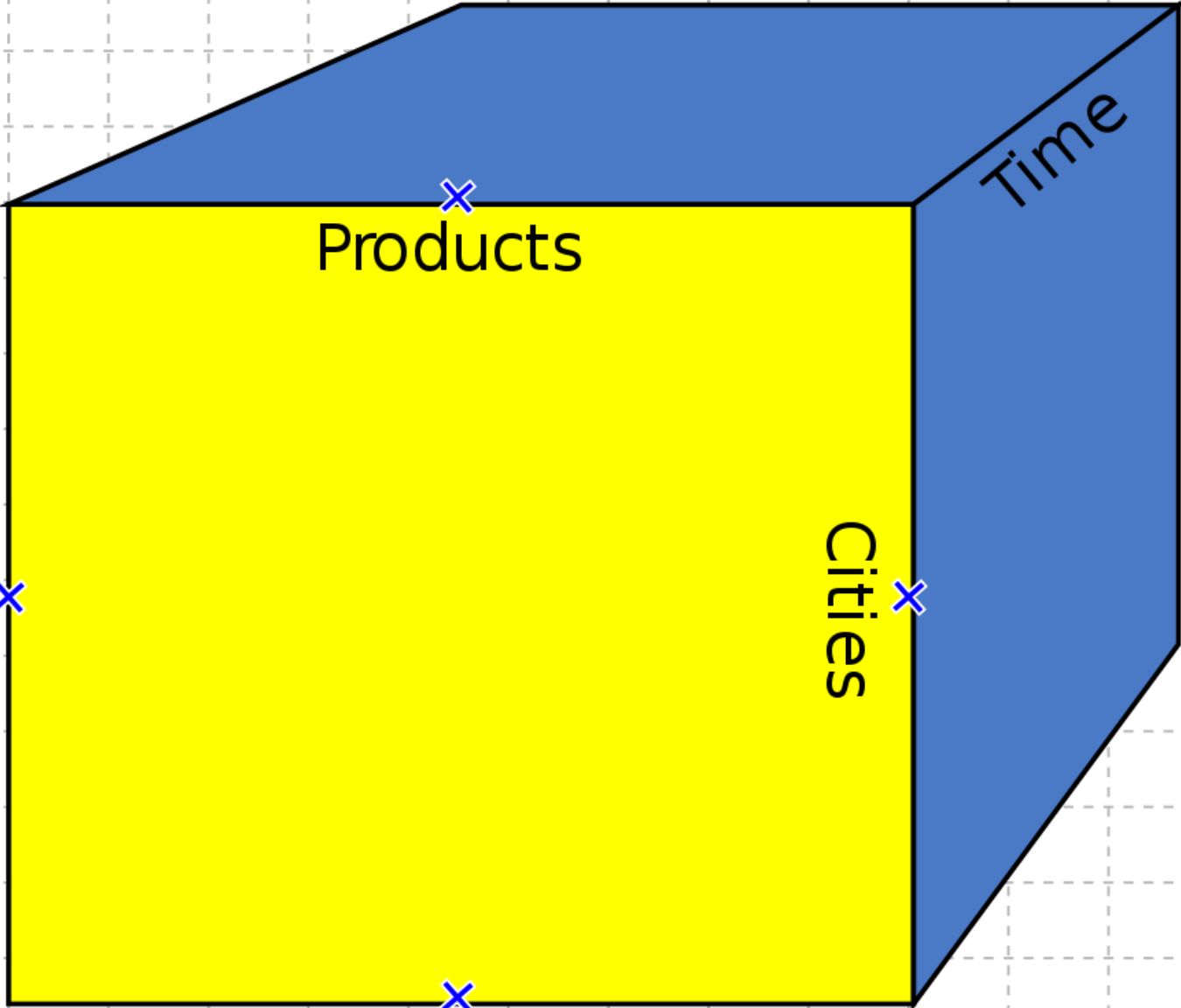
- Παραδοσιακό μοντέλο
  - OLTP
  - OLAP
- Big Data
- Share nothing architecture - Data sharding
- Συνέπεια δεδομένων - Θεωρήματα CAP & PACELC
- Μετάπτωση, ενοποίηση και διαλειτουργικότητα δεδομένων

# Online Transaction Processing (OLTP)

- Κεντρική έννοια: transaction / δοσοληψία
- Δύο οπτικές:
  - ACID transactions
    - Atomicity, Consistency, Isolation, Durability
    - Commit/rollback
  - Business/financial transactions
    - Μεταφορά ποσών μεταξύ λογαριασμών
    - Αγορές, πωλήσεις, κλπ.
- Στην πράξη: χρησιμοποιούμε τα ACID transactions για να υλοποιήσουμε τα business/financial transactions

# Online Analytical Processing (OLAP)

- Μέρος του Business Intelligence (BI)
- Reports
- Multi-dimensional Analysis (MDA)
- Hypercubes



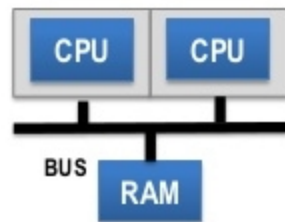
# Big Data

- Ανάλυση δεδομένων πέραν της παραδοσιακής OLAP προσέγγισης
- Volume, Variety, Velocity
- Παράλληλη επεξεργασία
- Τεχνικές τύπου MapReduce
- Προβλέψεις, στοχαστικά μοντέλα, μηχανική μάθηση, νευρωνικά δίκτυα, κλπ.

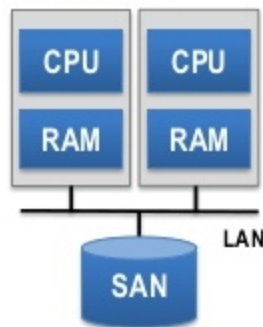
# MapReduce

```
function sum(a, b) { return a+b; }  
function getAge(person) { return currentYear - person.birthYear; }  
let records = //list of person records  
let ageSum = records.map(getAge).reduce(sum, 0)  
return ageSum / records.length; //avg age
```

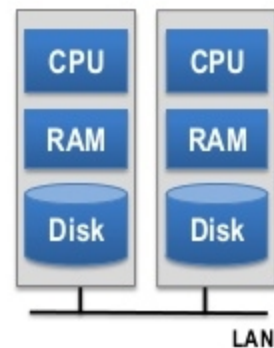
# Share nothing architecture



Shared RAM



Shared Disk



Shared Nothing



# Χαρακτηριστικά

- Κάθε κόμβος είναι ανεξάρτητος και αυτοτελής.
- No single point of contention (δεν διαμοιράζονται πόροι, π.χ. μνήμη ή δίσκος).
- Sharding: οριζόντια επιμέρηση των δεδομένων.
- Οριζόντια κλιμάκωση (horizontal scalability) - απλή προσθήκη κόμβων.
- Η αρχιτεκτονική πολλών NoSQL συστημάτων.

# Το Θεώρημα CAP

Σε περίπτωση αστοχίας δικτύου (P), θα έχουμε είτε συνέπεια των δεδομένων (C) είτε διαθεσιμότητα της εφαρμογής (A), αλλά όχι και τα δύο.

```
if (P) { A or C }
```

# Το Θεώρημα PACELC

Επέκταση του CAP

Αν δεν υπάρχει P, θα έχουμε είτε συνέπεια των δεδομένων (C) είτε την ελάχιστη δυνατή καθυστέρηση (L), αλλά όχι και τα δύο.

```
if (P) { A or C }  
else   { L or C }
```

# Κατηγοριοποίηση κατακεμημένων συστημάτων

PA/EL

```
if P then A else L //BASE Systems
```

PC/EC

```
if P then C else C //ACID databases
```

PC/EL

```
if P then C else L
```

# BASE Systems

- Basically Available, Soft state, Eventual consistency
- Eventually consistent: Όταν πάψουν οι ενημερώσεις σε μια εγγραφή, τελικά (eventually) όλες οι αναγνώσεις της εγγραφής αυτής θα επιστρέψουν την πιο πρόσφατη ενημέρωση.
- Replica convergence (σύγκλιση αντιγράφων)
- PA/EL (Επιλέγουν αυξημένη διαθεσιμότητα & μείωση καθυστέρησης αντί για συνέπεια)

# Μετάπτωση δεδομένων

- Η διαδικασία επιλογής, προετοιμασίας, εξαγωγής, μετασχηματισμού και μόνιμης μεταφοράς των δεδομένων από ένα σύστημα σε άλλο.
- Απαιτείται συνήθως επικύρωση των δεδομένων που μεταφέρθηκαν (για πληρότητα).
- Το ζητούμενο είναι να αυτοματοποιήσουμε τη διαδικασία όσο γίνεται περισσότερο.

# Είδη μετάπτωσης

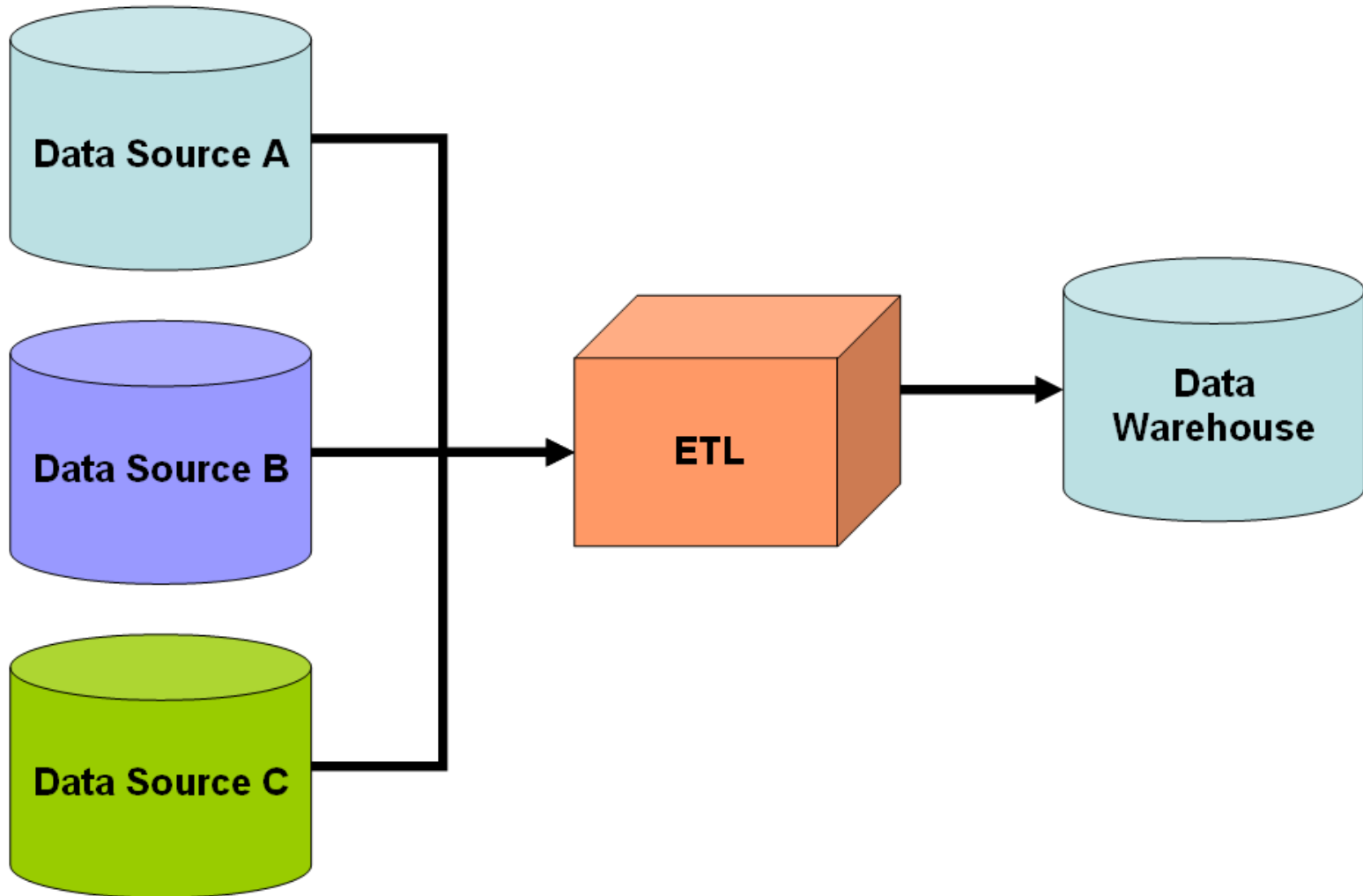
- Storage migration: αλλαγή του συστήματος αποθήκευσης (π.χ. δίσκοι)
- Database migration: αλλαγή του συστήματος βάσεων δεδομένων (π.χ. από το X RDBMS στο Y)
- Application migration: αλλαγή του πληροφοριακού συστήματος (π.χ. από το παλιό ERP στο καινούριο)
- Business process migration: αλλαγή της επιχειρησιακής διαδικασίας (π.χ. business process re-engineering, AS-IS vs TO-BE διαγράμματα, σε επόμενη διάλεξη)

# Data integration

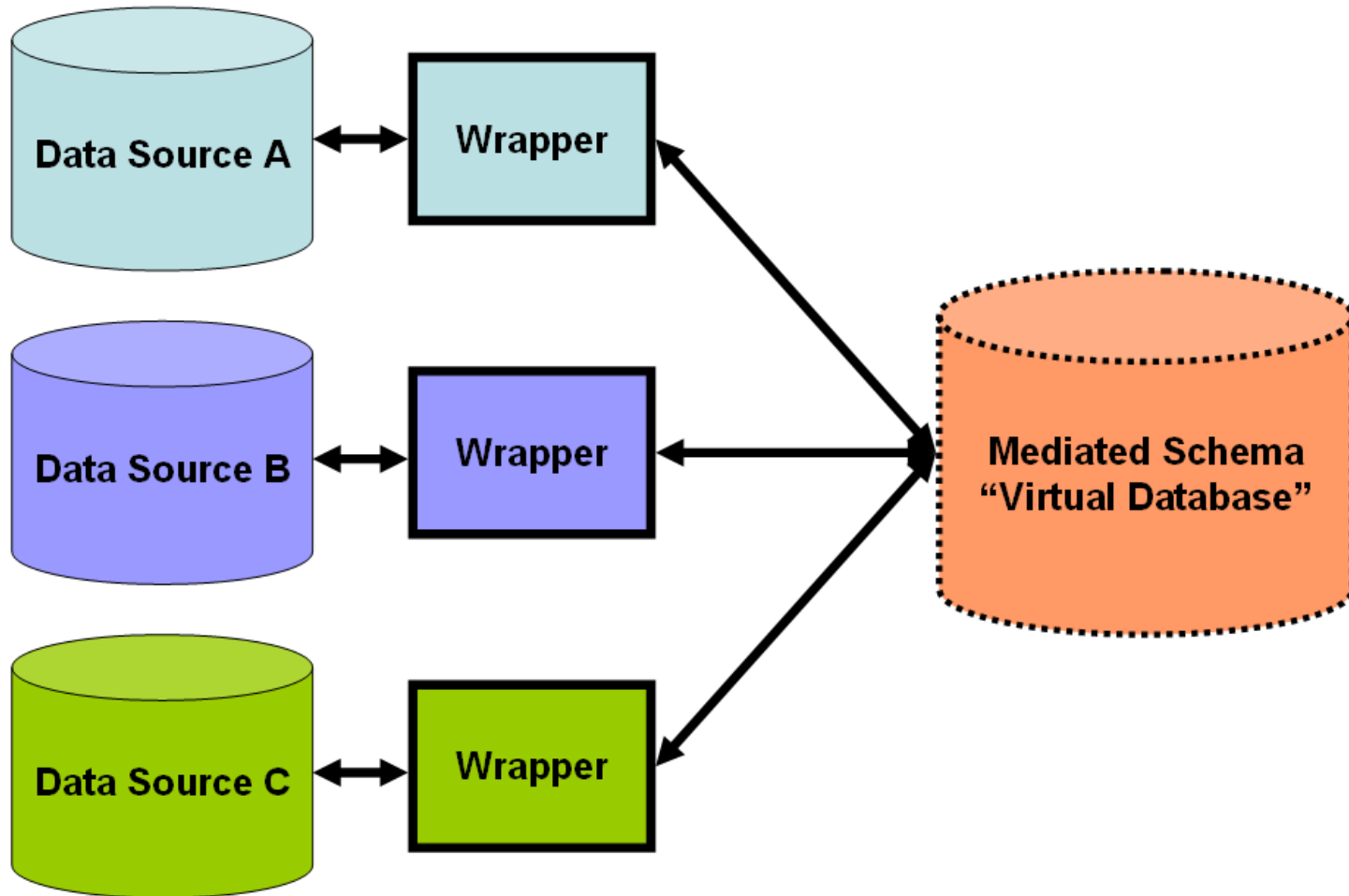
- Η ενοποίηση δεδομένων που βρίσκονται σε διαφορετικές πηγές και η παρουσίασή τους στους χρήστες σαν να βρίσκονταν αυτά σε μία
- Data warehousing (ETL), Data wrappers/bridges (Virtual Database)



# Extract, Transform, Load (ETL)



# Virtual database



# Data interoperability

- Οι διεπαφές (προγραμματιστικές) ενός συστήματος είναι πλήρως κατανοητές και μπορούν να "συνδεθούν" με άλλα συστήματα, στο παρόν ή στο μέλλον, χωρίς περιορισμούς.
- Ad-hoc / custom
- Διεθνή πρότυπα

# Παράδειγμα

- Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH)
- Διεθνές πρότυπο διαλειτουργικότητας για ψηφιακές βιβλιοθήκες και αποθετήρια (μετα)δεδομένων
- XML

# Ανοικτότητα

- Open Source
- Open Access
- Open Data
- Open APIs