

# Κανονικοποίηση βάσεων δεδομένων

1<sup>η</sup>, 2<sup>η</sup> και 3<sup>η</sup> κανονική μορφή

Αθανάσιος Σταυρακούδης

<http://stavrakoudis.econ.uoi.gr>

Άνοιξη 2014

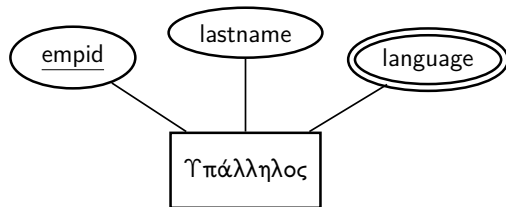


# Περιεχόμενα

- 1 Πρώτη κανονική μορφή
- 2 Συναρτησιακές εξαρτήσεις
- 3 Δεύτερη κανονική μορφή
- 4 Τρίτη κανονική μορφή



# Υπάλληλοι και γλώσσες



Υπάλληλοι μιας εταιρείας με πλειότερη ιδιότητα των ξένων γλωσσών που μιλούν. Οι υπόλοιπες ιδιότητες έχουν απαλειφθεί για λόγους απλότητας.



# Υπάλληλοι και γλώσσες σε πίνακα

## Δημιουργία του πίνακα υπαλλήλων

```
1 CREATE TABLE employees (  
2     empid      INTEGER NOT NULL  
3     lastname   VARCHAR(30) NOT NULL,  
4     language   VARCHAR(100),  
5     CONSTRAINT pk PRIMARY KEY (empid)  
6 );
```



# Υπάλληλοι και γλώσσες σε πίνακα

## Δημιουργία του πίνακα υπαλλήλων

```
1 CREATE TABLE employees (  
2     empid      INTEGER NOT NULL  
3     lastname   VARCHAR(30) NOT NULL,  
4     language   VARCHAR(100),  
5     CONSTRAINT pk PRIMARY KEY (empid)  
6 );
```

## Περιεχόμενα του πίνακα υπαλλήλων

empid	lastname	language
102	Διαμαντίδης	Αγγλικά, Γερμανικά
109	Αθανασίου	Αγγλικά
153	Αλεβιζάτου	Γαλλικά, Αγγλικά, Ρώσικα
172	Βλάσσης	Γαλλικά
189	Αγγελίνας	Γερμανικά, Αγγλικά



# Προβληματισμοί

- **SELECT:** Είναι εύκολο να γραφούν όλα τα πιθανά (ή τουλάχιστον τα περισσότερα) ερωτήματα επιλογής;
- **INSERT:** Μπορούν νέα γεγονότα να εισαχθούν εύκολα στους πίνακες της βάσης, χωρίς να απαιτούνται δύσκολοι και περίπλοκοι χειρισμοί;
- **DELETE:** Είναι εύκολη η διαγραφή δεδομένων, χωρίς βέβαια να χαθεί πολύτιμη πληροφορία;
- **UPDATE:** Μπορεί η τροποποίηση των δεδομένων της βάσης να γίνει με απλό τρόπο χωρίς πολύπλοκους χειρισμούς, και χωρίς να επηρεαστούν από τις αλλαγές δεδομένα που πρέπει να μείνουν όπως πριν;



# Αποτυχία ερωτημάτων επιλογής

σΓλώσσα='Αγγλικά' Υπάλληλος

```
1 SELECT *  
2 FROM employees  
3 WHERE language = 'English'
```

- Θα βρει την Αθανασίου που μιλάει μόνο Αγγλικά
- Δε θα βρει τον Αγγελίνα που μιλάει Αγγλικά και Γερμανικά

Πιθανή λύση:

```
1 SELECT *  
2 FROM employees  
3 WHERE language like '*English*'
```



# Πρόβλημα ενημέρωσης

Ας υποθέσουμε πως μετά τα Γαλλικά:

<b>Επώνυμο</b>	<b>Γλώσσες</b>
----------------	----------------

Βλάσσης	Γαλλικά
---------	---------

Ο κ. Παπαδόπουλος μαθαίνει και Αγγλικά:

<b>Επώνυμο</b>	<b>Γλώσσες</b>
----------------	----------------

Βλάσσης	Γαλλικά, Αγγλικά
---------	------------------

Πως θα γίνει η ενημέρωση της βάσης δεδομένων;

```
UPDATE employees
```

```
SET language = 'Γαλλικά, Αγγλικά'
```

```
WHERE name = 'Βλάσσης'
```

**Πρόβλημα:** Για να προσθέσουμε τη γνώση Αγγλικών στον Βλάσση χρειάζεται να ξέρουμε πως ξέρει ήδη Γαλλικά!





# Υπάλληλοι και γλώσσες - μια άλλη προσέγγιση

Επώνυμο	Αγγλικά	Γερμανικά	Γαλλικά	Ρώσικα
Διαμαντίδης	NAI	NAI	OXI	OXI
Αθανασίου	NAI	OXI	OXI	OXI
Αλεβιζάτου	OXI	OXI	OXI	OXI
Μαροπούλου	NAI	OXI	NAI	NAI
Βλάσσης	OXI	OXI	NAI	OXI
Αγγελίνας	NAI	NAI	OXI	OXI



# Τα νέα προβλήματα

**Αποθήκευση:** Απαιτείται αποθηκευτικός χώρος για δεδομένα που δεν χρειάζεται να αποθηκευτούν.

**Επιλογή:** Υπάρχει σχετικό πλεονέκτημα. Για παράδειγμα, μπορεί εύκολα να βρεθούν οι υπάλληλοι που μιλούν αγγλικά:

$$\pi_{\text{Επώνυμο}}(\sigma_{\text{Αγγλικά}='ΝΑΙ'}(\Upsilon\acute{\pi}\alpha\lambda\lambda\eta\lambda\omicron\iota))$$

**Εισαγωγή:** Ενώ η επιλογή δεδομένων μπορεί στοιχειωδώς να δουλέψει σωστά, η εισαγωγή είναι αρκετά προβληματική

**Διαγραφή:** Όπως και στην εισαγωγή, η διαγραφή στην απλή περίπτωση είναι τροποποίηση

**Τροποποίηση:** Η τροποποίηση, σε αντίθεση με τις προηγούμενες περιπτώσεις είναι απλή

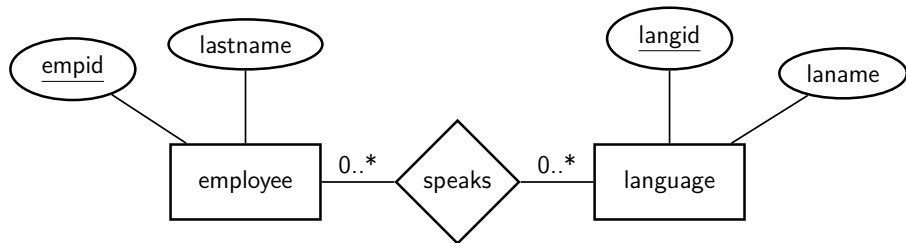


# Πρώτη κανονική μορφή

Επώνυμο	Γλώσσα
Διαμαντίδης	Αγγλικά
Διαμαντίδης	Γερμανικά
Αθανασίου	Αγγλικά
Αλεβιζάτου	Γαλλικά
Αλεβιζάτου	Αγγλικά
Αλεβιζάτου	Ρώσικα
Παπαδόπουλος	Γαλλικά
Αγγελίνας	Γερμανικά
Αγγελίνας	Αγγλικά



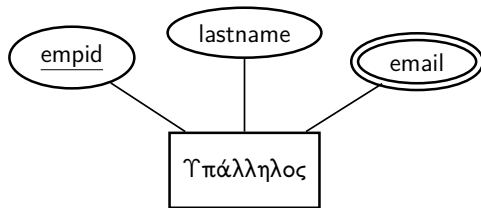
# Υπάλληλοι και γλώσσες, 1η ΚΜ



Η πλειότιμη ιδιότητα μετατρέπεται σε συσχέτιση  
**«πολλά-προς-πολλά»**.



# Λύστε αυτό το πρόβλημα



Υπάλληλοι μιας εταιρείας με πλειότιμη ιδιότητα των email επικοινωνίας.

Δώστε το σχεσιακό μοντέλο.

Οι υπάλληλοι μπορεί να έχουν από 0 μέχρι 2 email επικοινωνίας.



# Περιεχόμενα

- 1 Πρώτη κανονική μορφή
- 2 **Συναρτησιακές εξαρτήσεις**
- 3 Δεύτερη κανονική μορφή
- 4 Τρίτη κανονική μορφή



# Συναρτησιακή εξάρτηση

Ποδοσφαιριστής	Ομάδα	Πρωτάθλημα
Μαλντίνι	Μίλαν	Ιταλικό
Ανρύ	Άρσεναλ	Αγγλικό
Κάρλος	Ρεάλ Μαδρίτης	Ισπανικό
Μέσι	Μπαρτσελόνα	Ισπανικό
Τζέραλντ	Λίβερπουλ	Αγγλικό
Ροναλντίνιο	Μπαρτσελόνα	Ισπανικό
Τζιλαρντίνιο	Μίλαν	Ιταλικό

Ομάδα → Πρωτάθλημα



# Προσοχή στις συναρτησιακές εξαρτήσεις

Φοιτητής	Δήμος	Νομός
Βασιλείου	Αγρινίου	Αιτωλοακαρνανίας
Μαρονίτης	Αθήνας	Αττικής
Γεροδήμου	Αμπελοκήπων	Θεσσαλονίκης
Αντωνιάδης	Καβάλας	Καβάλας
Μακρίδου	Καλαμαριάς	Θεσσαλονίκης
Φιλίππου	Αθήνας	Αττικής
Γεωργιάδης	Αμπελοκήπων	Αττικής

Δήμος → Νομός





# Τετριμμένες συναρτησιακές εξαρτήσεις

Έστω σχέση  $r$  με σχήμα  $R = A, X, Y$  όπου  $A$  είναι πρωτεύον κλειδί, τότε:

$$A \rightarrow X$$

$$A \rightarrow Y$$

$$A \rightarrow XY$$

$$A \rightarrow AX$$

$$A \rightarrow AY$$

$$A \rightarrow AXY$$

$$A \rightarrow AXY$$

$$AX \rightarrow X$$

$$AX \rightarrow Y$$

$$AX \rightarrow XY$$

$$AX \rightarrow AY$$

$$AX \rightarrow AXY$$

$$AY \rightarrow X$$

$$AY \rightarrow Y$$

$$AY \rightarrow XY$$

$$AY \rightarrow AX$$

$$AY \rightarrow AXY$$

$$AXY \rightarrow A$$

$$AXY \rightarrow X$$

$$AXY \rightarrow Y$$

$$AXY \rightarrow AX$$

$$AXY \rightarrow AY$$

$$AXY \rightarrow XY$$



# Αξιιώματα του Armstrong

- ❶ **Ανακλαστικότητα.** Αν  $\alpha$  είναι σύνολα γνωρισμάτων  $\beta$  και  $\beta \subseteq \alpha$ , τότε  $\alpha \rightarrow \beta$ .
- ❷ **Επαύξηση.** Αν ισχύει  $\alpha \rightarrow \beta$  και  $\gamma$  είναι σύνολο γνωρισμάτων, τότε  $\alpha\gamma \rightarrow \beta\gamma$
- ❸ **Μεταβατικότητα.** Αν ισχύουν  $\alpha \rightarrow \beta$  και  $\beta \rightarrow \gamma$  τότε ισχύει  $\alpha \rightarrow \gamma$ .



# Περισσότεροι κανόνες

- ❶ **Αυτοκαθορισμός:**  $\alpha \rightarrow \alpha$
- ❷ **Ανακλαστικότητα:**  $\beta \subseteq \alpha \models \alpha \rightarrow \beta$
- ❸ **Ένωση:**  $\{\alpha \rightarrow \beta, \alpha \rightarrow \gamma\} \models \alpha \rightarrow \beta\gamma$
- ❹ **Διάσπαση:**  $\alpha \rightarrow \beta\gamma \models \alpha \rightarrow \beta, \alpha \rightarrow \gamma$
- ❺ **Σύνθεση:**  $\{\alpha \rightarrow \beta, \gamma \rightarrow \delta\} \models \alpha\beta \rightarrow \gamma\delta$
- ❻ **Ψευδομεταβατικότητα:**  $\{\alpha \rightarrow \beta, \beta\gamma \rightarrow \delta\} \models \alpha\gamma \rightarrow \delta$
- ❼ **Γενική ενοποίηση του Darwen:**  
 $\{\alpha \rightarrow \beta, \gamma \rightarrow \delta\} \models \alpha \cup (\gamma - \beta) \rightarrow \beta\delta$



# Μη αναγώγιμα σύνολα εξαρτήσεων

Για την καλή λειτουργία μιας βάσης δεδομένων είναι καλό να ξέρουμε το ελάχιστο σύνολο εξαρτήσεων. Ένα σύνολο συναρτησιακών εξαρτήσεων  $F$  λέγεται μη αναγώγιμο, ή ελάχιστο αν:

- 1 Το εξαρτώμενο (δεξιό) μέλος κάθε συναρτησιακής εξάρτησης αποτελείται από ένα μόνο γνώρισμα, δηλαδή δεν είναι σύνολο γνωρισμάτων.
- 2 Για κάθε συναρτησιακή εξάρτηση, το ορίζον μέλος είναι το μέγιστο δυνατό. Για παράδειγμα, αν  $\alpha, \beta$  και  $\gamma$  είναι γνωρίσματα με  $\gamma \subset \beta$ , η συναρτησιακή εξάρτηση  $\beta \rightarrow \alpha$  δεν μπορεί να αντικατασταθεί από την  $\gamma \rightarrow \alpha$  και να έχουμε ένα νέο ισοδύναμο σύνολο συναρτησιακών συναρτήσεων.



# Πλήρης συναρτησιακή εξάρτηση

Έστω η συναρτησιακή εξάρτηση  $QU \rightarrow Z$ . Τότε:

**Μερική** Η συναρτησιακή εξάρτηση λέγεται μερική όταν ισχύει είτε  $Q \rightarrow Z$ , είτε  $U \rightarrow Z$ . Δηλαδή είναι δυνατόν να αφαιρεθεί μέρος των γνωρισμάτων από το ορίζον (αριστερό) μέλος της συναρτησιακής εξάρτησης, και να παραμένει σε ισχύ.

**Πλήρης** Η συναρτησιακή συνάρτηση λέγεται πλήρης όταν είναι αδύνατη η ισχύς της, μετά την αφαίρεση κάποιων γνωρισμάτων από το ορίζον μέλος. Δηλαδή, η  $QU \rightarrow Z$  είναι πλήρης, όταν δεν ισχύει ούτε  $Q \rightarrow Z$ , ούτε  $U \rightarrow Z$ .



# Περιεχόμενα

- 1 Πρώτη κανονική μορφή
- 2 Συναρτησιακές εξαρτήσεις
- 3 Δεύτερη κανονική μορφή
- 4 Τρίτη κανονική μορφή



## Σύνθετο πρωτεύον κλειδί και συναρτησιακή εξάρτηση

title	actor	director
Blade Runner	Harrison Ford	Ridley Scott
Blade Runner	Sean Young	Ridley Scott
Blade Runner	Daryl Hannah	Ridley Scott
Gladiator	Harrison Ford	Ridley Scott
Gladiator	Connie Nielsen	Ridley Scott
Paris Texas	Sam Berry	Wim Wenders

Το πρωτεύον κλειδί είναι σύνθετο και αποτελείται από το συνδυασμό των γνωρισμάτων *title*, *actor*. Στη σχέση *films* ισχύουν οι συναρτησιακές εξαρτήσεις:

$$\{\text{title, actor}\} \rightarrow \text{director}$$

title  $\rightarrow$  director



# Διάσπαση για μετατροπή σε 2η κανονική μορφή

film\_actor  $\bowtie$  film\_direct

title	actor
Blade Runner	Harrison Ford
Blade Runner	Sean Young
Blade Runner	Daryl Hannah
Gladiator	Harrison Ford
Gladiator	Connie Nielsen
Paris, Texas	Sam Berry
title	director
Blade Runner	Ridley Scott
Gladiator	Ridley Scott
Paris, Texas	Wim Wenders





# Δημιουργία σχήματος

```
CREATE TABLE film_direct  
(  
    title      CHAR(25) NOT NULL,  
    director   CHAR(25) NOT NULL,  
    CONSTRAINT pk PRIMARY KEY(title)  
);
```

```
CREATE TABLE film_actor  
(  
    title CHAR(25) NOT NULL,  
    actor CHAR(25) NOT NULL,  
    CONSTRAINT pk PRIMARY KEY(title , actor),  
    CONSTRAINT fk_title FOREIGN KEY (title)  
        REFERENCES film_direct (title)  
);
```



# Μεταφορά δεδομένων

```
INSERT INTO film_direct (title , director)  
  SELECT DISTINCT title , director  
  FROM films;
```

```
INSERT INTO film_actor (title , actor)  
  SELECT title , actor  
  FROM films;
```



# Περιεχόμενα

- 1 Πρώτη κανονική μορφή
- 2 Συναρτησιακές εξαρτήσεις
- 3 Δεύτερη κανονική μορφή
- 4 Τρίτη κανονική μορφή



# Τρίτη κανονική μορφή

Η τρίτη κανονική μορφή αναφέρεται στις συναρτησιακές εξαρτήσεις γνωρισμάτων μιας σχέσης που δεν ανήκουν στο πρωτεύον κλειδί.

Το σχήμα  $R$  μιας σχέσης  $r$  είναι σε 3η ΚΜ όταν είναι σε 2η ΚΜ και για κάθε συναρτησιακή εξάρτηση:

$$X \rightarrow Y$$

όπου  $X, Y \subset R$ , το  $X$  είναι είτε υπερκλειδί, είτε πρωτεύον κλειδί της  $r$ .

Σε αντίθετη περίπτωση, αν δηλαδή το ορίζον μέλος μιας συναρτησιακής εξάρτησης δεν είναι πρωτεύον κλειδί, ή υπερκλειδί της σχέσης, τότε το σχήμα δεν βρίσκεται σε τρίτη κανονική μορφή.



# Κανονικοποίηση με διάσπαση

Θεωρείστε ένα σχεσιακό σχήμα  $R = \{X, Y, Z\}$  με πρωτεύον κλειδί το  $X$ . Θεωρείστε επίσης πως ισχύει  $Y \rightarrow Z$ . Τότε το  $R$  δεν είναι σε 3η ΚΜ. Στην ουσία το  $R$  φέρει μια μεταβατική συναρτησιακή εξάρτηση:

$$\begin{array}{lcl} X & \rightarrow & Y \\ Y & \rightarrow & Z \end{array} \quad \} \Rightarrow X \rightarrow Z$$

- 1 Το εξαρτημένο μέλος της συναρτησιακής εξάρτησης  $Y \rightarrow Z$ , δηλαδή το  $Z$ , απομακρύνετε από το αρχικό σχήμα  $R$ . Έτσι  $R1 = \{X, Y\}$  με πρωτεύον κλειδί το  $X$ , όπως και στην περίπτωση του σχήματος  $R$ .
- 2 Τα μέλη της συναρτησιακής εξάρτησης  $Y \rightarrow Z$ , τόσο το  $Y$ , όσο και το εξαρτώμενο  $Z$ , σχηματίζουν ένα νέο σχήμα  $R2 = \{Y, Z\}$  με πρωτεύον κλειδί το  $Y$ , δηλαδή το  $Y$  ορίζον μέλος της συναρτησιακής εξάρτησης.
- 3 Η αρχική σχέση μπορεί να αναδημιουργηθεί από τη σύζευξη:



# Παράδειγμα μη τρίτης ΚΜ

id	name	island	region
1045	Απέραντο Γαλάζιο	Ρόδος	Δωδεκαννήσου
632	Παράδεισος	Κάλυμνος	Δωδεκαννήσου
3019	Παράδεισος	Κως	Δωδεκαννήσου
8397	Ηλιοβασίλεμα	Νάξος	Κυκλάδων
894	Άμμος	Νάξος	Κυκλάδων
4283	Έλενα	Σκιάθος	Μαγνησίας



# Κανονικοποίηση με διάσπαση

Η σχέση δεν είναι σε 3η ΚΜ, γιατί υπάρχει η συναρτησιακή εξάρτηση:

*island*  $\rightarrow$  *region*

Ο κανόνας κανονικοποίησης επιβάλλει τη διάσπασή της σε δύο απλούστερες σχέσεις:

- Το εξαρτώμενο μέλος της συναρτησιακής εξάρτησης:

*island*  $\rightarrow$  *region*

δηλαδή το *region*, αφαιρείται από τη σχέση *hotels*.

- Τα μέλη της συναρτησιακής εξάρτησης σχηματίζουν μια νέα σχέση, πχ με όνομα *islands* με πρωτεύον κλειδί το ορίζον μέλος, δηλαδή το *islands*.
- Η αρχική σχέση *hotels* μπορεί να αναδημιουργηθεί από τη σύζευξη:

*hotels* = *places*  $\bowtie$  *islands*



# SQL για το κανονικοποιημένο σχήμα

```
CREATE TABLE islands
(
    island CHAR(20) NOT NULL,
    region CHAR(20),
    CONSTRAINT pk PRIMARY KEY (island)
);
```

```
CREATE TABLE places
(
    id      INTEGER NOT NULL,
    name    CHAR(25) NOT NULL,
    island  CHAR(20) NOT NULL,
    CONSTRAINT pk PRIMARY KEY (id),
    CONSTRAINT fk_island FOREIGN KEY (island)
        REFERENCES islands (island)
);
```





# Μεταφορά δεδομένων

```
INSERT INTO islands (island , region)  
SELECT DISTINCT island , region  
FROM hotels;
```

```
INSERT INTO places (id , name, island)  
SELECT id , name, island  
FROM hotels;
```



# Νέο σχεσιακό σχήμα

id	name	island
1045	Απέραντο Γαλάζιο	Ρόδος
632	Παράδεισος	Κάλυμνος
3019	Παράδεισος	Κως
8397	Ηλιοβασίλεμα	Νάξος
894	Άμμος	Νάξος
4283	Έλενα	Σκιάθος

island	region
Ρόδος	Δωδεκαννήσου
Κάλυμνος	Δωδεκαννήσου
Κως	Δωδεκαννήσου
Νάξος	Κυκλάδων
Σκιάθος	Μαγνησίας



# Απόδοση στα ερωτήματα

Για παράδειγμα, για να βρεθεί το νησί και ο νομός του ξενοδοχείου με κωδικό 632, στην μη κανονικοποιημένη βάση θα εκτελεστεί το ερώτημα:

$$\pi_{island, region}(\sigma_{id=632}(hotels))$$

```
SELECT island , region  
FROM hotels  
WHERE id = 632;
```

Αντίθετα, στην κανονικοποιημένη βάση, θα χρειαστεί σύζευξη:

$$\pi_{island, region}(\sigma_{id=632}(islands \bowtie places))$$

```
SELECT island , region  
FROM places p INNER JOIN islands i  
ON p.island = i.island
```



Σας ευχαριστώ για την προσοχή σας

Είμαι στη διάθεσή σας για σχόλια, απορίες και ερωτήσεις

