



EPL342 –Databases

Lecture 21: Functional Dependencies and Normalization Informal Design Guidelines (Chapter 14.1, Elmasri-Navathe 7ED)

Demetris Zeinalipour

<http://www.cs.ucy.ac.cy/courses/EPL342>

Εισαγωγή στην Κανονικοποίηση (Normalization)



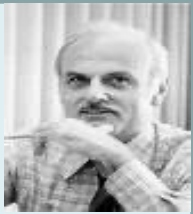
- Η Σχεδίαση του Σχεσιακού Σχήματος μιας βάσης δεδομένων μέχρι στιγμής γινόταν βάσει της **κοινής λογικής** του **σχεδιαστή** και όχι βάσει **αυστηρών κανόνων**.
 - Τα παραγόμενα σχήματα μπορούσαν να κριθούν **υποκειμενικά** ως προς την **Χρηστότητα (Goodness)** τους αλλά όχι **αντικειμενικά**.
- Παρόλο που αναφερθήκαμε στο παρελθόν σε κάποιες γενικές οδηγίες χαρακτηρισμού της **Χρηστότητας (Goodness)** ενός Σχεσιακού Σχήματος, π.χ.,
 - Να **ΜΗΝ** υπάρχει **επανάληψη πληροφορίας**
 - Να **ΜΗΝ** υπάρχουν **πολλά NULLs** σε γνωρίσματα.
 - Να **ΜΗΝ** απαιτούνται **πολλές συνενώσεις** για να καταλήξουμε σε μια οντότητα.

Αυτές οι οδηγίες δεν ήταν τυπικές (δηλ, δεν στηρίζονταν σε κάποιο μαθηματικό υπόβαθρο οπότε ήταν υποκειμενικές)!²¹⁻³

Εισαγωγή στην Κανονικοποίηση (Normalization)



- **Κανονικοποίηση (Normalization)** : Η συστηματική (τυπική) διαδικασία διασφάλισης ότι το σχήμα μιας βάσης δεδομένων τις πιο κάτω πρόνοιες:
 - Είναι κατάλληλο για εκτέλεση γενικού είδους επερωτήσεων
 - Είναι ελεύθερο από ορισμένα μη-επιθυμητά χαρακτηριστικά (ανωμαλίες) εισαγωγής, ενημέρωσης, και διαγραφής τα οποία μπορεί να οδηγήσουν σε απώλεια δεδομένων
- Η Κανονικοποίηση ορίζεται το 1970 από τον **Edgar Frank "Ted" Codd**, τον δημιουργό του Σχεσιακού Μοντέλου.
 - Αυτό για να επιτευχθεί ευκολότερα η μετάβαση των τότε βάσεων δεδομένων (ιεραρχικό και δικτυακό μοντέλο) στο **σχεσιακό μοντέλο**.
- Ο Codd ορίζει μέχρι το 1974 διάφορα επίπεδα Κανονικοποίησης (**1NF, 2NF, 3NF και BCNF**)
 - ...τα οποία μπορούν να ακολουθηθούν βήμα-βήμα για προσδιορισμό ενός καλού σχεσιακού σχήματος.



Εισαγωγή στην Κανονικοποίηση (Normalization)



- Η Κανονικοποίηση στηρίζεται στην έννοια των **συναρτησιακών εξαρτήσεων (functional dependencies)**, τα οποία αποτελούν γενίκευση της έννοιας του κλειδιού.
 - Π.χ., $SSN \rightarrow \{Name, NumberGrade, LetterGrade, Dno\}$
 $NumberGrade \rightarrow LetterGrade$
 - Αυτά θα μελετηθούν εκτενώς στην επόμενη διάλεξη.
- **Προτού δούμε** τις πιο πάνω έννοιες θα επιχειρήσουμε να **συνοψίσουμε με άτυπο τρόπο** κάποιες **γενικές οδηγίες** για το **σχεδιασμό** ενός καλού σχήματος μιας σχεσιακής βάσης.

Πλεονάζουσα Πληροφορία (Προβλήματα)



- **Πλεονάζουσα Πληροφορία σε Σχέσεις:**
Δεδομένα τα οποία αποθηκεύονται **επαναληπτικά (redundantly)** προκαλούν τα ακόλουθα προβλήματα:

A) Σπαταλούν Χώρο Αποθήκευσης, εάν και αυτό το πρόβλημα είναι μικρό σε σχέση με το επόμενο

B) Προκαλούν προβλήματα με τις ενημερώσεις (update anomalies)

- Ανωμαλίες Εισαγωγής (Insertion anomalies)
- Ανωμαλίες Διαγραφής (Deletion anomalies)
- Ανωμαλίες Τροποποίησης (Modification anomalies)

Ας δούμε κάποια παραδείγματα...

Πλεονάζουσα Πληροφορία (Σπατάλη Χώρου)



A) Σπατάλη Χώρου Αποθήκευσης (Redundancy) σε δυο σχέσεις Emp_Dept και Emp_Proj

- Και οι δυο περιπτώσεις είναι αποτέλεσμα άκομψου σχεδιασμού του σχήματος των πινάκων

EMP_DEPT

Ename	Ssn	Bdate	Address	Dnumber	Dname	Dmgr_ssn
Smith, John B.	123456789	1965-01-09	731 Fondren, Houston, TX	5	Research	333445555
Wong, Franklin T.	333445555	1955-12-08	638 Voss, Houston, TX	5	Research	333445555
Zelaya, Alicia J.	999887777	1968-07-19	3321 Castle, Spring, TX	4	Administration	987654321
Wallace, Jennifer S.	987654321	1941-06-20	291 Berry, Bellaire, TX	4	Administration	987654321
Narayan, Ramesh K.	666884444	1962-09-15	975 FireOak, Humble, TX	5	Research	333445555
English, Joyce A.	453453453	1972-07-31	5631 Rice, Houston, TX	5	Research	333445555
Jabbar, Ahmad V.	987987987	1969-03-29	980 Dallas, Houston, TX	4	Administration	987654321
Borg, James E.	888665555	1937-11-10	450 Stone, Houston, TX	1	Headquarters	888665555

Redundancy

EMP_PROJ

Ssn	Pnumber	Hours	Ename	Pname	Plocation
123456789	1	32.5	Smith, John B.	ProductX	Bellaire
123456789	2	7.5	Smith, John B.	ProductY	Sugarland
666884444	3	40.0	Narayan, Ramesh K.	ProductZ	Houston
453453453	1	20.0	English, Joyce A.	ProductX	Bellaire
453453453	2	20.0	English, Joyce A.	ProductY	Sugarland
333445555	2	10.0	Wong, Franklin T.	ProductY	Sugarland
333445555	3	10.0	Wong, Franklin T.	ProductZ	Houston
333445555	10	10.0	Wong, Franklin T.	Computerization	Stafford
333445555	20	10.0	Wong, Franklin T.	Reorganization	Houston
999887777	30	30.0	Zelaya, Alicia J.	Newbenefits	Stafford
999887777	10	10.0	Zelaya, Alicia J.	Computerization	Stafford
987987987	10	35.0	Jabbar, Ahmad V.	Computerization	Stafford
987987987	30	5.0	Jabbar, Ahmad V.	Newbenefits	Stafford
987654321	30	20.0	Wallace, Jennifer S.	Newbenefits	Stafford
987654321	20	15.0	Wallace, Jennifer S.	Reorganization	Houston
888665555	20	Null	Borg, James E.	Reorganization	Houston

Redundancy

Redundancy

Ανωμαλίες Ενημερώσεων (Insert Anomaly)



- **Ανωμαλία Εισαγωγής (Insert Anomaly)**
 - **Πρόβλημα 1:** Για εισαγωγή **νέας πλειάδας** πρέπει να βεβαιωθούμε ότι το (Dname, Dmgr_ssn) είναι συνεπές με τις προηγούμενες εγγραφές.
 - Π.χ., Dname="Research", Dmgr_ssn=987654321 **λάθος. Γιατί;**
 - **Πρόβλημα 2:** Δεν μπορούμε να εισάγουμε τα στοιχεία ενός Department το οποίο δεν έχει ακόμη employees!
 - Δεν γίνεται ούτε εάν βάλουμε **NULL** στα ανύπαρκτα πεδία. **Γιατί;**

EMP_DEPT

Ename	Ssn	Bdate	Address	Dnumber	Dname	Dmgr_ssn
Smith, John B.	123456789	1965-01-09	731 Fondren, Houston, TX	5	Research	333445555
Wong, Franklin T.	333445555	1955-12-08	638 Voss, Houston, TX	5	Research	333445555
Zelaya, Alicia J.	999887777	1968-07-19	3321 Castle, Spring, TX	4	Administration	987654321
Wallace, Jennifer S.	987654321	1941-06-20	291 Berry, Bellaire, TX	4	Administration	987654321
Narayan, Ramesh K.	666884444	1962-09-15	975 FireOak, Humble, TX	5	Research	333445555
English, Joyce A.	453453453	1972-07-31	5631 Rice, Houston, TX	5	Research	333445555
Jabbar, Ahmad V.	987987987	1969-03-29	980 Dallas, Houston, TX	4	Administration	987654321
Borg, James E.	888665555	1937-11-10	450 Stone, Houston, TX	1	Headquarters	888665555

Ανωμαλίες Ενημερώσεων (Insert Anomaly)



- Το πιο κάτω στιγμιότυπο των Employee-Department δεν υποφέρει από τις ανωμαλίες εισαγωγής της προηγούμενης σελίδας

EMPLOYEE				
Ename	Ssn	Bdate	Address	Dnumber
Smith, John B.	123456789	1965-01-09	731 Fondren, Houston, TX	5
Wong, Franklin T.	333445555	1955-12-08	638 Voss, Houston, TX	5
Zelaya, Alicia J.	999887777	1968-07-19	3321 Castle, Spring, TX	4
Wallace, Jennifer S.	987654321	1941-06-20	291Berry, Bellaire, TX	4
Narayan, Ramesh K.	666884444	1962-09-15	975 Fire Oak, Humble, TX	5
English, Joyce A.	453453453	1972-07-31	5631 Rice, Houston, TX	5
Jabbar, Ahmad V.	987987987	1969-03-29	980 Dallas, Houston, TX	4
Borg, James E.	888665555	1937-11-10	450 Stone, Houston, TX	1

DEPARTMENT		
Dname	Dnumber	Dmgr_ssn
Research	5	333445555
Administration	4	987654321
Headquarters	1	888665555

A) Δεν χρειάζεται έλεγχος συνέπειας για εισαγωγή στο Department ή Employee

B) Μπορούμε να εισάγουμε Departments, Employees ανεξάρτητα το ένα από την ύπαρξη του άλλου.

Ανωμαλίες Ενημερώσεων (Delete Anomaly)



- **Ανωμαλία Διαγραφής (Delete Anomaly)**
 - Πρόβλημα: Εάν σβήσουμε την τελευταία πλειάδα που περιέχει τα στοιχεία ενός Department, τότε θα χαθούν όλες οι πληροφορίες του εν λόγω Department.
 - Π.χ., DELETE FROM EMP_DEPT WHERE SSN=888665555
 - Το πρόβλημα αυτό **ΔΕΝ** θα προκύψει στο **στιγμιότυπο της προηγούμενης διαφάνειας**
 - όπου τα στοιχεία ενός department συνεχίζουν να διατηρούνται στον πίνακα Department

EMP_DEPT

Ename	Ssn	Bdate	Address	Dnumber	Dname	Dmgr_ssn
Smith, John B.	123456789	1965-01-09	731 Fondren, Houston, TX	5	Research	333445555
Wong, Franklin T.	333445555	1955-12-08	638 Voss, Houston, TX	5	Research	333445555
Zelaya, Alicia J.	999887777	1968-07-19	3321 Castle, Spring, TX	4	Administration	987654321
Wallace, Jennifer S.	987654321	1941-06-20	291 Berry, Bellaire, TX	4	Administration	987654321
Narayan, Ramesh K.	666884444	1962-09-15	975 FireOak, Humble, TX	5	Research	333445555
English, Joyce A.	453453453	1972-07-31	5631 Rice, Houston, TX	5	Research	333445555
Jabbar, Ahmad V.	987987987	1969-03-29	980 Dallas, Houston, TX	4	Administration	987654321
Borg, James E.	888665555	1937-11-10	450 Stone, Houston, TX	1	Headquarters	888665555

Redundancy

Ανωμαλίες Ενημερώσεων (Update Anomaly)



- **Ανωμαλία Ενημέρωσης (Update Anomaly)**

- **Πρόβλημα:** Εάν ενημερώσουμε το γνώρισμα ενός στοιχείου το οποίο επαναλαμβάνεται τότε πρέπει να ενημερώσουμε **ΟΛΕΣ** τις εμφανίσεις του.
- Αυτό το πρόβλημα είναι το μικρότερο από τα αναφερόμενα διότι λύνεται με μια έκφραση SQL .
 - Π.χ., UPDATE EMP_DEPT SET Dname='RES' WHERE Dname='Research'
 - ...εάν και το query θα απαιτεί περισσότερο χρόνο από ότι πρέπει

EMP_DEPT

Ename	Ssn	Bdate	Address	Dnumber	Dname	Dmgr_ssn
Smith, John B.	123456789	1965-01-09	731 Fondren, Houston, TX	5	Research	333445555
Wong, Franklin T.	333445555	1955-12-08	638 Voss, Houston, TX	5	Research	333445555
Zelaya, Alicia J.	999887777	1968-07-19	3321 Castle, Spring, TX	4	Administration	987654321
Wallace, Jennifer S.	987654321	1941-06-20	291 Berry, Bellaire, TX	4	Administration	987654321
Narayan, Ramesh K.	666884444	1962-09-15	975 FireOak, Humble, TX	5	Research	333445555
English, Joyce A.	453453453	1972-07-31	5631 Rice, Houston, TX	5	Research	333445555
Jabbar, Ahmad V.	987987987	1969-03-29	980 Dallas, Houston, TX	4	Administration	987654321
Borg, James E.	888665555	1937-11-10	450 Stone, Houston, TX	1	Headquarters	888665555