



Επισκόπηση Σχεδίασης Σχεσιακών Βάσεων Δεδομένων

Γιάννης Θεοδωρίδης

InfoLab, Τμήμα Πληροφορικής, Πανεπιστήμιο Πειραιά

<http://infolab.cs.unipi.gr>

version: Oct.2009

Περιεχόμενα



- Εισαγωγή
 - Μοντελοποίηση, Σχεδιασμός και Ανάπτυξη ΒΔ
- Σχεδιασμός ΒΔ
 - Εννοιολογικός σχεδιασμός - το Μοντέλο Οντοτήτων-Συσχετίσεων (Entity-Relationship)
 - Λογικός σχεδιασμός - το Σχεσιακό (Relational) Μοντέλο
 - Σχεσιακή Άλγεβρα
 - Κανονικές μορφές

Περιεχόμενα



- **Εισαγωγή**
 - **Μοντελοποίηση, Σχεδιασμός και Ανάπτυξη ΒΔ**
- **Σχεδιασμός ΒΔ**
 - Εννοιολογικός σχεδιασμός - το Μοντέλο Οντοτήτων-Συσχετίσεων (Entity-Relationship)
 - Λογικός σχεδιασμός - το Σχεσιακό (Relational) Μοντέλο
 - Σχεσιακή Άλγεβρα
 - Κανονικές μορφές

Ορισμοί (1)



- **Δεδομένα** (data)

Γνωστά γεγονότα / περιστατικά που καταγράφονται και έχουν νόημα
- **Βάση Δεδομένων** (database)

Μία ολοκληρωμένη συλλογή δεδομένων που συσχετίζονται

 - αποθηκεύεται (συνήθως) σε δευτερεύουσα μνήμη

Η Βάση Δεδομένων αναπαριστά ένα υποσύνολο των στοιχείων που μπορούμε να συλλέξουμε για τον μικρόκοσμο (πεδίο αναφοράς) μιας εφαρμογής.

Ορισμοί (2)



- **Σύστημα Διαχείρισης Βάσεων Δεδομένων (ΣΔΒΔ)**

(Database Management System - DBMS)

Μια αυτοτελής συλλογή από τμήματα λογισμικού (προγράμματα) για τη δημιουργία, επεξεργασία και συντήρηση βάσεων δεδομένων.

- **Σύστημα Βάσεων Δεδομένων** (database system)

Το λογισμικό (ΣΔΒΔ) μαζί με τη Βάση Δεδομένων.

Εργαλεία μοντέλων δεδομένων



- Αφαιρετικά:

Πραγματικός κόσμος $\Rightarrow$ Δεδομένα $\Rightarrow$ Μοντέλο

- Πέρα από την καταγραφή των δεδομένων, χρειάζεται να τα **οργανώσουμε / ερμηνεύσουμε / συσχετίσουμε**
- Ένα ΜΟΝΤΕΛΟ ΔΕΔΟΜΕΝΩΝ είναι ένα εργαλείο που επιτρέπει αυτή την ΑΦΑΙΡΕΣΗ
(ανεξαρτήτως των συγκεκριμένων δεδομένων)

Ερωτήσεις κατά τη μοντελοποίηση των δεδομένων



- Πώς πρέπει να **μοντελοποιήσουμε** αυτό τον μικρόκοσμο?
- Ποιες είναι οι βασικές **συνιστώσες** του?
- Τι γνωρίζουμε **για αυτές** ?
- Πώς πρέπει να **οργανωθούν**?
- Πώς **συσχετίζονται** ?
- Ποιες πληροφορίες έχουν **ενδιαφέρον** ? Ποιες είναι **άσχετες** ?
- Σε **ποιον** έχουν ενδιαφέρον ? Για **ποιο σκοπό** ?
- Πόσο **συχνά** ?
- Κάτω από ποιες **συνθήκες** ?

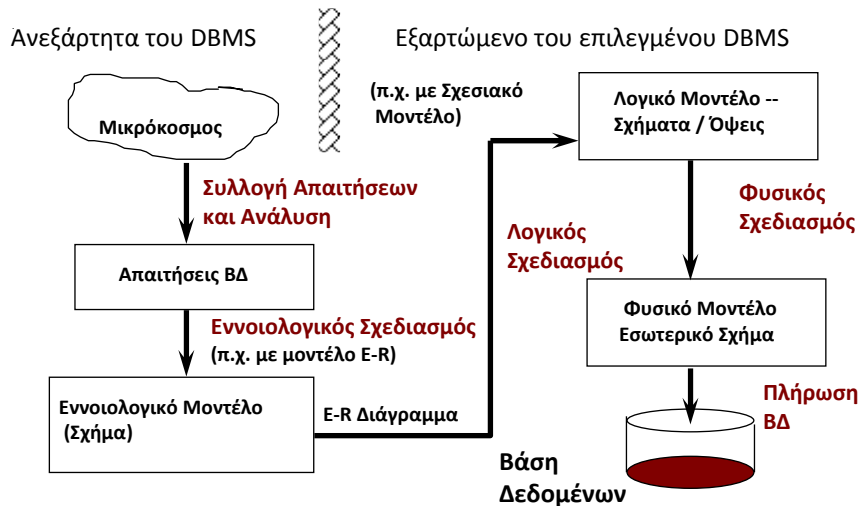
...

Σχεδίαση ΒΔ



- Η διαδικασία της σχεδίασης ΒΔ
 - Μια επίπονη διαδικασία μετατροπής της **γνώσης** που έχουμε για έναν μικρόκοσμο σε μια ΒΔ που θα την διαχειρίζεται ένα ΣΔΒΔ
 - Για εννοιολογική διευκόλυνση της όλης διαδικασίας, αυτή διακρίνεται σε διαδοχικές αυτοτελείς διαδικασίες (**φάσεις**), κάθε μια των οποίων δημιουργεί ενδιάμεσες περιγραφές.
 - **Συλλογή και Ανάλυση Απαιτήσεων** (Requirements Collection and Analysis)
 - **Εννοιολογικός Σχεδιασμός ΒΔ** (Conceptual Database Design)
 - **Λογικός Σχεδιασμός ΒΔ** (Logical Database Design)
 - **Φυσικός Σχεδιασμός ΒΔ** (Physical Database Design)

Διαδικασία σχεδίασης ΒΔ (1)



9

ΠΑ.ΠΕΙ. - Γιάννης Θεοδωρίδης

Διαδικασία σχεδίασης ΒΔ (2)



■ Συλλογή και Ανάλυση Απαιτήσεων

Η πρώτη φάση ανάπτυξης της ΒΔ είναι απαραίτητη για την «ανακάλυψη» των απαιτήσεων για τη ΒΔ από περιγραφές του πραγματικού μικρόκοσμου.

Περιγραφές Μικρόκοσμου ➡ Απαιτήσεις της Βάσης

- Η διαδικασία είναι «δύσκολη» (εννοιολογικά) και απαιτεί εμπειρία και επιχειρηματική γνώση (συνήθως στηρίζεται σε συνεντεύξεις, ασαφείς και ημιτελείς περιγραφές, κλπ.)

10

ΠΑ.ΠΕΙ. - Γιάννης Θεοδωρίδης

Διαδικασία σχεδίασης ΒΔ (3)



■ Σχεδιασμός

- Σχεδιασμός είναι η διαδικασία δημιουργίας του **σχήματος** (schema) της ΒΔ χρησιμοποιώντας ένα επιλεγμένο Μοντέλο (σε κάποιο επίπεδο της όλης διαδικασίας)
- Υπάρχουν 3 αυτοτελείς σχεδιασμοί που δημιουργούν σχήματα (περιγραφές) της Βάσης.

Εννοιολογικός Σχεδιασμός $\Rightarrow$ (Conceptual Data Model)

Λογικός Σχεδιασμός $\Rightarrow$ (Logical Schema & External Schemas - Views)

Φυσικός Σχεδιασμός $\Rightarrow$ (Internal Schema)

Εννοιολογικός Σχεδιασμός



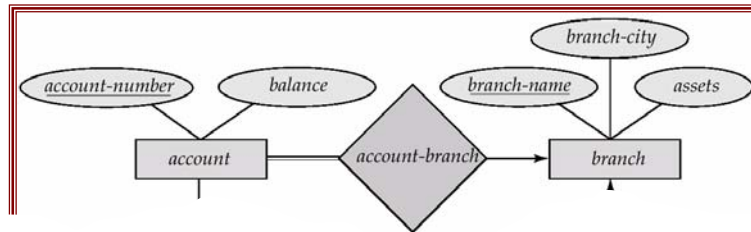
(Conceptual Modeling)

- Προσπάθεια για ξεκαθάρισμα των εννοιών
- Εύρεση και καταγραφή των **Οντοτήτων** της ΒΔ, των μεταξύ τους **Συσχετίσεων** καθώς και των Εννοιολογικών Κανόνων (semantic rules) που πρέπει να ικανοποιούν

■ ΣΤΟΧΟΣ:

Μια αφαιρετική, αλλά πλήρης περιγραφή του τμήματος του μικρόκοσμου που θα αναπαρασταθεί στην βάση δεδομένων. Αυτή η περιγραφή γίνεται με τη χρήση μιας ημι-τυπικής σημειογραφίας / συμβολισμού.

Παράδειγμα εννοιολογικού σχεδιασμού



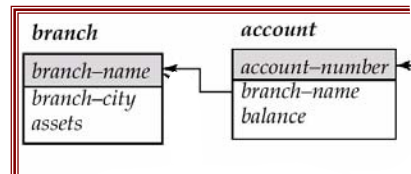
- Μοντέλο οντοτήτων-συσχετίσεων (Entity-Relationship model – ER)
 - Τα υποκαταστήματα (branches) ανοίγουν λογαριασμούς καταθέσεων (accounts)
 - για τους καταθετικούς λογαριασμούς, ενδιαφερόμαστε να γνωρίζουμε κωδικό και υπόλοιπο
 - για τα υποκαταστήματα, ενδιαφερόμαστε να γνωρίζουμε όνομα, έδρα και αποθεματικό

Λογικός Σχεδιασμός



- Η διαδικασία μετατροπής ενός εννοιολογικού μοντέλου (διαισθητικής περιγραφής) σε **τυπικά σχήματα** εκφρασμένα στο επιλεγέν (υποστηριζόμενο από το ΣΔΒΔ) μοντέλο δεδομένο (π.χ. Σχεσιακό Μοντέλο)
 - Μερικά σχήματα, αν είναι καλά σχεδιασμένα, κάνουν ευκολότερη την υποβολή των ερωτήσεων.
 - Μερικά σχήματα, αν είναι καλά σχεδιασμένα, αποφεύγουν τους πλεονασμούς, τα προβλήματα σε αλλαγές των δεδομένων, κλπ.
- ➔ Θέματα ακεραιότητας ΒΔ, κανονικοποίησης κλπ.

Παράδειγμα λογικού σχεδιασμού



- Το σχεσιακό μοντέλο (relational model) μας οδηγεί σε 2 σχέσεις-πίνακες:
 - υποκαταστήματα
 - καταθετικοί λογαριασμοί

Φυσικός Σχεδιασμός



- Η διαδικασία προδιαγραφών των **αρχείων** που υλοποιούν τον λογικό σχεδιασμό και των οργανώσεων αυτών, καθώς και των **δομών ευρετηρίων** που αποτελούν το εσωτερικό σχήμα.

ΣΤΟΧΟΙ:

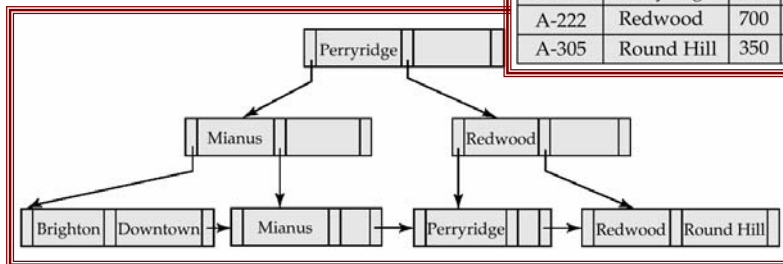
- Απόδοση
- Υποστήριξη του Βελτιστοποιητή Συστήματος

Παράδειγμα φυσικού σχεδιασμού



- ένα διατεταγμένο αρχείο για τον πίνακα account, μαζί με
- ένα ευρετήριο B⁺-δέντρο πάνω στο πεδίο branch-name

A-217	Brighton	750	
A-101	Downtown	500	
A-110	Downtown	600	
A-215	Mianus	700	
A-102	Perryridge	400	
A-201	Perryridge	900	
A-218	Perryridge	700	
A-222	Redwood	700	
A-305	Round Hill	350	



Περιεχόμενα



- Εισαγωγή
 - Μοντελοποίηση, Σχεδιασμός και Ανάπτυξη ΒΔ
- Σχεδιασμός ΒΔ
 - Εννοιολογικός σχεδιασμός - το Μοντέλο Οντοτήτων-Συσχετίσεων (Entity-Relationship)
 - Λογικός σχεδιασμός - το Σχεσιακό (Relational) Μοντέλο
 - Σχεσιακή Άλγεβρα
 - Κανονικές μορφές

Μοντέλο Οντοτήτων-Συσχετίσεων



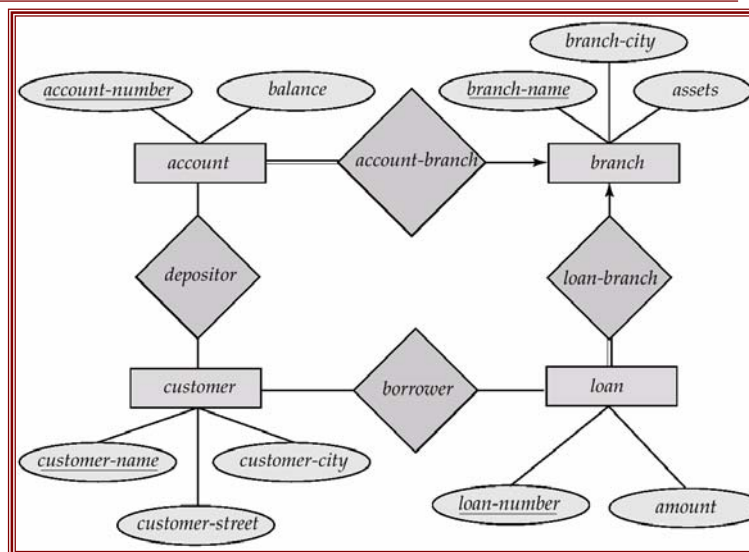
- Το E-R (Entity-Relationship) μοντέλο προτάθηκε το 1976 από τον **Peter P.-S. Chen**

**“An Entity-Relationship model – toward a unified view of data”,
ACM TODS, 1(1):9-36, January 1976**

ήταν ο «νικητής» μεταξύ αρκετών υποψηφίων (τη δεκαετία του 80) και έγινε αποδεκτό από τους ερευνητές και την αγορά για

- την απλότητα,
 - το σαφήνεια και
 - το γραφικό συμβολισμό που έχει.
- Το E-R μοντέλο είναι ένα μάλλον **διαισθητικό** μοντέλο που επιδιώκει να προσδιορίσει **αφαιρετικά** τις πληροφορίες που μια ΒΔ αποθηκεύει και οργανώνει.

Παράδειγμα E-R



Συστατικά του E-R Μοντέλου



- Υπάρχουν δυο βασικές εννοιολογικές έννοιες:
 - **Οντότητες** - αντικείμενα που υπάρχουν (ή πιστεύεται ότι υπάρχουν) και μπορούν να αναπαρασταθούν στην ΒΔ
 - Π.χ. ΠΕΛΑΤΗΣ, ΥΠΟΚΑΤΑΣΤΗΜΑ
 - **Συσχετίσεις** - (ειδικά) αντικείμενα που αντιστοιχούν δύο ή περισσότερες ξεχωριστές οντότητες με ένα συγκεκριμένο νόημα
 - Π.χ. ένας ΠΕΛΑΤΗΣ είναι κάτοχος κάποιου ΛΟΓΑΡΙΣΜΟΥ
- Οι Οντότητες (αλλά και οι Συσχετίσεις) μπορούν να έχουν **χαρακτηριστικά** (attributes) που είναι ιδιότητες που τις χαρακτηρίζουν. Είδη χαρακτηριστικών:
 - **Απλά**, π.χ., υπόλοιπο λογαριασμού
 - **Σύνθετα**, π.χ., διεύθυνση = {Δρόμος, Αριθμός, ΤΚ, Πόλη, Χώρα}
 - **Πλειότιμα**, π.χ., τηλέφωνο(-α) ενός πελάτη

Περιεχόμενα

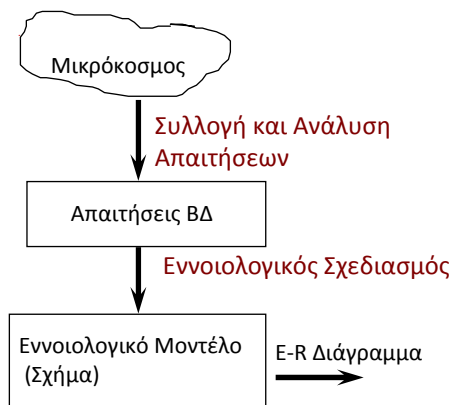


- Εισαγωγή
 - Μοντελοποίηση, Σχεδιασμός και Ανάπτυξη ΒΔ
- Σχεδιασμός ΒΔ
 - Εννοιολογικός σχεδιασμός - το Μοντέλο Οντοτήτων-Συσχετίσεων (Entity-Relationship)
 - **Λογικός σχεδιασμός - το Σχεσιακό (Relational) Μοντέλο**
 - Σχεσιακή Άλγεβρα
 - Κανονικές μορφές

Επισκόπηση διαδικασίας σχεδίασης ΒΔ



Ανεξάρτητα του ΣΔΒΔ



Μετά την χρήση ενός Εννοιολογικού Μοντέλου Δεδομένων, όπως το E-R, ερχόμαστε σε ένα σημείο όπου αρχίζει η εξάρτηση από το επιλεγέν ΣΔΒΔ και πρέπει να γίνει η **μετατροπή** της «διαισθητικής» περιγραφής σε «τυπική», που να μπορεί να εκτελεστεί από το ΣΔΒΔ.

Λογικός Σχεδιασμός



Σε αυτό το σημείο **ΕΠΙΛΕΓΟΥΜΕ** το Μοντέλο Δεδομένων που υποστηρίζεται από το ΣΔΒΔ που έχουμε διαθέσιμο (ή θεωρούμε κατάλληλο).

Ο βασικός υποψήφιος είναι το **Σχεσιακό Μοντέλο**

(άλλοι πιθανοί υποψήφιοι είναι το Object-Oriented, το Object-Relational, και το ημι-δομημένο XML μοντέλο)

Εννοιολογικό Σχήμα → E-R διάγραμμα

Εξάρτηση από το επιλεγέν ΣΔΒΔ

Σχεσιακό Μοντέλο → Λογικό Σχήμα και Εξωτερικές Όψεις

Λογικός Σχεδιασμός ΒΔ

Εισαγωγή στο Σχεσιακό Μοντέλο (1)



- Προτάθηκε το 1970 από τον **E.F. Codd**

**“A relational model of data for large shared data banks”,
Communications of the ACM, 13(6):377-387, June 1970**

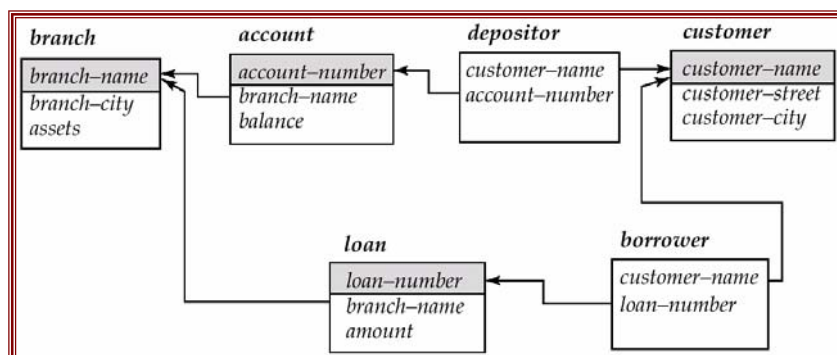
ως Θεωρία για Μοντέλα Δεδομένων

- Έδωσε έναυσμα για πολλές ερευνητικές προσπάθειες και κατέληξε να είναι το πλέον δημοφιλές μοντέλο.
 - Σήμερα, η συντριπτική πλειοψηφία των ΣΔΒΔ είναι Σχεσιακά και διατίθενται σε όλες τις υπολογιστικές πλατφόρμες.

Εισαγωγή στο Σχεσιακό Μοντέλο (2)



- Ένα **Σχεσιακό Σχήμα** (Relational schema) είναι ένα Σύνολο Σχέσεων (πινάκων)



Εισαγωγή στο Σχεσιακό Μοντέλο (3)



- Ένα **πεδίο τιμών** (domain) D είναι ένα σύνολο τιμών

$$D = \{d_1, d_2, \dots, d_n\}$$

- Ένα **χαρακτηριστικό** (attribute) A ονοματίζει μια (ενδιαφέρουσα) ιδιότητα σε μία Σχέση και παίρνει τιμές από το αντίστοιχο πεδίο τιμών $D(A)$.

Συμβολισμός: **$R.A$** ή **$R[A]$** , όπου R είναι το όνομα της Σχέσης

- Μια **πλειάδα** (tuple) t μιας Σχέσης $R(A_1, A_2, \dots, A_n)$ είναι μια (διατεταγμένη) λίστα τιμών $t = \langle v_1, v_2, \dots, v_n \rangle$, όπου κάθε τιμή v_i είναι ένα στοιχείο του πεδίου $D(A_i)$.

customer-name	customer-street	customer-city	χαρακτηριστικά (ή στήλες)
Jones	Main	Harrison	πλειάδες (ή γραμμές)
Smith	North	Rye	
Curry	North	Rye	
Lindsay	Park	Pittsfield	

customer

27

ΠΑ.ΠΕΙ. - Γιάννης Θεοδωρίδης

Εισαγωγή στο Σχεσιακό Μοντέλο (4)



- Η διάταξη των χαρακτηριστικών σε μια σχέση έχει σημασία
- Η διάταξη των πλειάδων σε μια σχέση δεν έχει σημασία
- Κάθε πλειάδα αποθηκεύεται μία φορά σε μια σχέση (η σχέση αποτελεί σύνολο πλειάδων)
- Μια τιμή μπορεί να παρουσιάζεται πολλές φορές σε μια στήλη και είναι ατομική (μη διασπασίμη)
- Συμβολισμός:
τιμή γνωρίσματος για μια πλειάδα t ,
 $t[A_i] = v_i$

account-number	branch-name	balance
A-101	Downtown	500
A-215	Mianus	700
A-102	Perryridge	400
A-305	Round Hill	350
A-201	Brighton	900
A-222	Redwood	700
A-217	Brighton	750

28

ΠΑ.ΠΕΙ. - Γιάννης Θεοδωρίδης

Εισαγωγή στο Σχεσιακό Μοντέλο (5)



- Υπάρχουν περιορισμοί που είναι **έμφυτοι** στο μοντέλο:
 - Περιορισμοί ακεραιότητας οντότητας (entity integrity)
 - Περιορισμοί αναφορικής ακεραιότητας (referential integrity)
- Υπάρχουν επίσης **ρητοί** περιορισμοί:
 - Περιορισμοί πεδίου τιμών
 - Περιορισμοί οριζόμενοι από το χρήστη (user-defined)

Εισαγωγή στο Σχεσιακό Μοντέλο (6)



- Εντοπίστε κάποιους έμφυτους και ρητούς περιορισμούς ...

account-number	branch-name	balance
A-101	Downtown	500
A-102	Perryridge	400
A-201	Brighton	900
A-215	Mianus	700
A-217	Brighton	750
A-222	Redwood	700
A-305	Round Hill	350

customer-name	account-number
Hayes	A-102
Johnson	A-101
Johnson	A-201
Jones	A-217
Lindsay	A-222
Smith	A-215
Turner	A-305

customer-name	customer-street	customer-city
Adams	Spring	Pittsfield
Brooks	Senator	Brooklyn
Curry	North	Rye
Glenn	Sand Hill	Woodside
Green	Walnut	Stamford
Hayes	Main	Harrison
Johnson	Alma	Palo Alto
Jones	Main	Harrison
Lindsay	Park	Pittsfield
Smith	North	Rye
Turner	Putnam	Stamford
Williams	Nassau	Princeton

Εισαγωγή στο Σχεσιακό Μοντέλο (7)



- Πράξεις στο ΣΜ
 - Οι πράξεις στο ΣΜ διαχωρίζονται σε ενημερώσεις και ανακτήσεις
 - Το σύνολο των πράξεων στο ΣΜ είναι **κλειστό**, δηλαδή οι πράξεις ορίζονται πάνω σε Σχέσεις και έχουν ως αποτέλεσμα νέες Σχέσεις
- Ενημερώσεις σε Σχέσεις
 - Εισαγωγή (INSERT) πλειάδας
 - Διαγραφή (DELETE) πλειάδας
 - Τροποποίηση (UPDATE) πλειάδας
- Οι περιορισμοί ακεραιότητας δεν πρέπει να παραβιάζονται με την εκτέλεση μιας πράξης ενημέρωσης !

Πράξεις ανάκτησης



- Υπάρχουν δύο βασικές "γλώσσες" (ομάδες πράξεων ΣΜ):
 - **Σχεσιακή Άλγεβρα** (relational algebra)
 - διαδικαστικού τύπου, προσδιορίζει **πώς** υπολογίζεται το αποτέλεσμα
 - **Σχεσιακός Λογισμός** (relational calculus)
 - δηλωτικού τύπου, προσδιορίζει **ποιες** ιδιότητες πρέπει να πληροί το αποτέλεσμα
- Κανένα σχεσιακό ΣΔΒΔ δεν υποστηρίζει τις δύο κατηγορίες γλωσσών στην πλήρη θεωρητική μορφή τους (για λόγους "ευκολίας χρήσης" κλπ.).
- Όμως, οι γλώσσες που έχουν ενσωματωθεί στα ΣΔΒΔ (π.χ. SQL) έχουν τις ρίζες τους είτε στην Σχεσιακή Άλγεβρα ή στον Σχεσιακό Λογισμό.

Επισκόπηση της Σχεσιακής Άλγεβρας (1)



- Η ΣΑ αποτελείται από ένα σύνολο τελεστών που ενεργούν σε Σχέσεις και έχουν σαν αποτέλεσμα Σχέσεις (η Άλγεβρα είναι κλειστή).
- Οι τελεστές, όπως και στην αριθμητική άλγεβρα, μπορεί να είναι **εμφωλευμένοι** (nested), εφόσον το αποτέλεσμα κάθε πράξης είναι μια Σχέση
- Υπάρχουν δύο είδη τελεστών / πράξεων:
 - κλασσικές πράξεις θεωρίας συνόλων
 - ένωση, τομή, διαφορά, καρτεσιανό γινόμενο, ...
 - εξειδικευμένες για βάσεις δεδομένων
 - προβολή, επιλογή, σύνδεση, ...

Επισκόπηση της Σχεσιακής Άλγεβρας (2)



- Κλασσικές πράξεις θεωρίας συνόλων:
 - **Ένωση**: Ένωσε όλες τις πλειάδες των δύο Σχέσεων σε μια Σχέση
$$R \cup S = \{ t \mid t \text{ is in } R \text{ or } t \text{ is in } S \}$$
 - **Τομή**: Βάλε όλες τις κοινές πλειάδες των δύο Σχέσεων σε μια Σχέση
$$R \cap S = \{ t \mid t \text{ is in } R \text{ and } t \text{ is in } S \}$$
 - **Διαφορά**: Επέλεξε τις πλειάδες της πρώτης Σχέσης, που δεν ανήκουν στη δεύτερη Σχέση
$$R - S = \{ t \mid t \text{ is in } R \text{ and } t \text{ is not in } S \}$$
 - **Καρτεσιανό Γινόμενο**: Συνδύασε τις πλειάδες της μιας Σχέσης με κάθε πλειάδα της άλλης
$$R \times S = \{ t \mid t \text{ is the concatenation of a Tuple in } R \text{ with a Tuple in } S \}$$

Επισκόπηση της Σχεσιακής Άλγεβρας (3)



- Η πράξη της **Επιλογής** (σ):
επιλογή από τις πλειάδες μιας
Σχέσης R αυτών που πληρούν
κάποια **συνθήκη c**

- η συνθήκη είναι μια λογική παρά-
σταση στα χαρακτηριστικά της R
- το αποτέλεσμα είναι μια Σχέση,
“οριζόντιο” υποσύνολο της R

account-number	branch-name	balance
A-101	Downtown	500
A-215	Mianus	700
A-102	Perryridge	400
A-305	Round Hill	350
A-201	Brighton	900
A-222	Redwood	700
A-217	Brighton	750

$$\sigma_c(R) = \{ t \mid t \text{ is in } r(R) \text{ and condition } c \text{ holds for } t \}$$

- Η πράξη της επιλογής είναι αντιμεταθετική και προσεταιριστική:
 - $\sigma_{c1}(\sigma_{c2}(R)) = \sigma_{c2}(\sigma_{c1}(R))$
 - $\sigma_{c1}(\sigma_{c2}(R)) = \sigma_{c1 \text{ AND } c2}(R) = \sigma_{c1, c2}(R)$
 - $\sigma_{c1}(\sigma_{c2}(\sigma_{c3}(R))) = \sigma_{c2}(\sigma_{c3}(\sigma_{c1}(R)))$

35

ΠΑ.ΠΕΙ. - Γιάννης Θεοδωρίδης

Επισκόπηση της Σχεσιακής Άλγεβρας (4)



- Η πράξη της **Προβολής** (π):
διατήρηση μερικών από τα χαρακτη-
ριστικά της R (που προσδιορίζονται
από μια λίστα L)

- το αποτέλεσμα είναι μια Σχέση,
“κάθετο” υποσύνολο της R

account-number	branch-name	balance
A-101	Downtown	500
A-215	Mianus	700
A-102	Perryridge	400
A-305	Round Hill	350
A-201	Brighton	900
A-222	Redwood	700
A-217	Brighton	750

$$\pi_L(R) = \{ t[L] \mid t \text{ is in } r(R) \text{ and } L \subset R \}$$

- Επειδή μια Σχέση είναι σύνολο, μπορεί να διαγραφούν και πλειάδες
στο τελικό αποτέλεσμα της προβολής, δηλαδή $CARD(\pi_L(R)) \leq CARD(R)$

36

ΠΑ.ΠΕΙ. - Γιάννης Θεοδωρίδης

Επισκόπηση της Σχεσιακής Άλγεβρας (5)



- Η πράξη της **Σύνδεσης** ($\bowtie$)
- Υπάρχουν πολλές μορφές σύνδεσης – όλες συνδυάζουν δύο Σχέσεις για την δημιουργία μιας τρίτης
 - (Θήτα-) σύνδεση ((theta) join)
 - ισοσύνδεση (equality join)
 - φυσική σύνδεση (natural join)
 - ημι-σύνδεση (semi-join)
 - εξωτερική σύνδεση (outer join)

Επισκόπηση της Σχεσιακής Άλγεβρας (6)



account-number	branch-name	balance
A-101	Downtown	500
A-215	Mianus	700
A-102	Perryridge	400
A-305	Round Hill	350
A-201	Brighton	900
A-222	Redwood	700
A-217	Brighton	750

customer-name	account-number
Hayes	A-102
Johnson	A-101
Johnson	A-201
Jones	A-217
Lindsay	A-222
Smith	A-215
Turner	A-305

customer-name	customer-street	customer-city
Adams	Spring	Pittsfield
Brooks	Senator	Brooklyn
Curry	North	Rye
Glenn	Sand Hill	Woodside
Green	Walnut	Stamford
Hayes	Main	Harrison
Johnson	Alma	Palo Alto
Jones	Main	Harrison
Lindsay	Park	Pittsfield
Smith	North	Rye
Turner	Putnam	Stamford
Williams	Nassau	Princeton

Επισκόπηση της Σχεσιακής Άλγεβρας (7)



- **(Θήτα-) Σύνδεση:** Συνδυασμός των πλειάδων δύο Σχέσεων που «ταιριάζουν» (δηλαδή, πληρούν κάποια Boolean συνθήκη θ) σε κάποια προσδιοριζόμενα χαρακτηριστικά

$$R \bowtie_{\theta} S$$

- Μια Θήτα-Σύνδεση είναι ισοδύναμη ενός Καρτεσιανού γινομένου που ακολουθείται από μια επιλογή με συνθήκη θ .

$$R \bowtie_{\theta} S = \sigma_{\theta} (R \times S)$$

- Η Σχέση-αποτέλεσμα έχει όλα τα χαρακτηριστικά της R και της S
- **Ισο-σύνδεση:** Συνδυασμός των πλειάδων δύο Σχέσεων που «ταιριάζουν» (έχουν την ίδια τιμή) σε κάποια χαρακτηριστικά
 - Ειδική περίπτωση της θ -Σύνδεσης όπου η συνθήκη θ είναι ισότητα.

Επισκόπηση της Σχεσιακής Άλγεβρας (8)



- **Φυσική σύνδεση:** Συνδυασμός των πλειάδων δύο Σχέσεων που «ταιριάζουν» (έχουν την ίδια τιμή) σε **όλα τα κοινά** χαρακτηριστικά

$$R \bowtie S$$

- Ειδική περίπτωση της ισο-σύνδεσης.
- Η Σχέση-αποτέλεσμα έχει τα κοινά χαρακτηριστικά μόνο μία φορά

- **Ημι-σύνδεση:** Επιλογή του υποσυνόλου μιας Σχέσης που συνδέεται με μια άλλη

$$R \bowtie_{\theta} S$$

- ισοδύναμη με μια σύνδεση ακολουθούμενη από μια προβολή.

$$R \bowtie_{\theta} S = \pi_L (R \bowtie_{\theta} S) \mid L: \text{τα χαρακτηριστικά της } R$$

account-number	branch-name	balance
A-101	Downtown	500
A-215	Mianus	700
A-102	Perryridge	400
A-305	Round Hill	350
A-201	Brighton	900
A-222	Redwood	700
A-217	Brighton	750

customer-name	account-number
Hayes	A-102
Johnson	A-101
Johnson	A-201
Jones	A-217
Lindsay	A-222
Smith	A-215
Turner	A-305

Επισκόπηση της Σχεσιακής Άλγεβρας (9)



- **Εξωτερική σύνδεση:** Συνδύασε πλειάδες δύο Σχέσεων που «ταιριάζουν», εμφανίζοντας όμως και τις μη «ταιριαστές» πλειάδες με τιμές null στα χαρακτηριστικά της άλλης Σχέσης

- Κίνητρο: Σε μια τυπική σύνδεση, οι μη «ταιριαστές» πλειάδες των δύο σχέσεων δεν παρουσιάζονται στο αποτέλεσμα

- Υπάρχουν παραλλαγές:

- **Left outer join** (παρουσιάζονται όλες οι πλειάδες στην R),
- **Right outer join** (παρουσιάζονται όλες οι πλειάδες στην S),
- **Full outer join** (παρουσιάζονται όλες οι πλειάδες στην R και την S)

customer-name	account-number
Hayes	A-102
Johnson	A-101
Johnson	A-201
Jones	A-217
Lindsay	A-222
Smith	A-215
Turner	A-305

customer-name	customer-street	customer-city
Adams	Spring	Pittsfield
Brooks	Senator	Brooklyn
Curry	North	Rye
Glenn	Sand Hill	Woodside
Green	Walnut	Stamford
Hayes	Main	Harrison
Johnson	Alma	Palo Alto
Jones	Main	Harrison
Lindsay	Park	Pittsfield
Smith	North	Rye
Turner	Putnam	Stamford
Williams	Nassau	Princeton

41

ΠΑ.ΠΕΙ. - Γιάννης Θεοδωρίδης

Επισκόπηση της Σχεσιακής Άλγεβρας (10)



- Το σύνολο των πράξεων {σ, π, ∪, −, ×} ονομάζεται **πλήρες σύνολο πράξεων** της ΣΑ.

- Οι υπόλοιπες πράξεις της ΣΑ μπορούν να περιγραφούν ως μια ακολουθία πράξεων από το σύνολο αυτό, π.χ.

$$R \bowtie_{\theta} S = \sigma_{\theta} (R \times S)$$

- Όμως, υπάρχουν και λειτουργίες/πράξεις που **δεν** υποστηρίζονται στη ΣΑ
- Ένα παράδειγμα: η **αναδρομική κλειστότητα** (transitive closure), μια αναδρομική συσχέτιση μεταξύ πλειάδων του ίδιου τύπου
 - π.χ. Υπάλληλος – Προϊστάμενος Υπάλληλος (employee – supervisor), Ανταλλακτικό – Τμήμα ανταλλακτικού (part – sub-part), κλπ.
 - Κλασικές Ερωτήσεις: Περιπλανώμενος Πωλητής (από πόλη σε πόλη μέσω πόλεων), Προϊστάμενοι σε όλα τα επίπεδα ενός Υπαλλήλου, ...

42

ΠΑ.ΠΕΙ. - Γιάννης Θεοδωρίδης

Συναρτησιακές εξαρτήσεις



- Μια **συναρτησιακή εξάρτηση** (functional dependency) $X \rightarrow Y$ ισχύει για τη Σχέση R εάν, για κάθε πλειάδα r της R:
 - $t_1 \in r, t_2 \in r, \pi_X(t_1) = \pi_X(t_2) \Rightarrow \pi_Y(t_1) = \pi_Y(t_2)$
 - δηλαδή, αν ταυτίζονται οι τιμές δύο πλειάδων στο χαρακτηριστικό (ή σύνολο χαρακτηριστικών) X θα πρέπει να ταυτίζονται και στο Y
- Εάν K είναι **υποψήφιο κλειδί** για την R τότε ισχύει ότι $K \rightarrow R$
- Αξιώματα του Armstrong (X, Y, Z είναι σύνολα χαρακτηριστικών)
 - **Ανακλαστική ιδιότητα** (reflexivity): if $X \supseteq Y$ then $X \rightarrow Y$
 - **Επαυξητική ιδιότητα** (augmentation): if $X \rightarrow Y$ then $XZ \rightarrow YZ$ for any Z
 - **Μεταβατική ιδιότητα** (transitivity): if $X \rightarrow Y$ and $Y \rightarrow Z$ then $X \rightarrow Z$

Κανονικές Μορφές (1)



- Ο ρόλος των συναρτησιακών εξαρτήσεων στην ανίχνευση πλεονασμών
 - Προφανώς, αν δεν υπάρχουν συναρτησιακές εξαρτήσεις μεταξύ των χαρακτηριστικών μιας σχέσης τότε δεν υπάρχει **πλεονασμός** (redundancy)
 - Αν αντίθετα ισχύει π.χ. $A \rightarrow B$, τότε οι πλειάδες με την ίδια τιμή στο A, υποχρεωτικά θα έχουν ίδια τιμή στο B.
 - Παράδειγμα πλεονασμού:

branch-name	branch-city	assets	customer-name	loan-number	amount
Downtown	Brooklyn	9000000	Jones	L-17	1000
Redwood	Palo Alto	2100000	Smith	L-23	2000
Perryridge	Horseneck	1700000	Hayes	L-15	1500
Downtown	Brooklyn	9000000	Jackson	L-14	1500

- Στόχος της κανονικοποίησης: να φέρουμε το σχήμα ΒΔ σε μια τέτοια **κανονική μορφή** (normal form -- BCNF, 3NF, ...) ώστε να αποφεύγονται ή τουλάχιστον να ελαχιστοποιούνται τα προβλήματα που μπορεί να προκύψουν λόγω πλεονασμού

Κανονικές Μορφές (2)



■ Boyce-Codd Normal Form (BCNF)

- Η Σχέση R είναι σε BCNF εάν για όλες τις συναρτησιακές εξαρτήσεις $X \rightarrow Y$, ισχύει ότι: X είναι υποψήφιο κλειδί της R
- Δηλαδή, δεν πρέπει ένα χαρακτηριστικό να έχει συναρτησιακή εξάρτηση από ένα άλλο χαρακτηριστικό εκτός από το κλειδί
- Με άλλα λόγια, οι μόνες συναρτησιακές εξαρτήσεις που επιτρέπονται είναι οι περιορισμοί κλειδιού

Κανονικές Μορφές (3)



■ Παράδειγμα αποσύνθεσης BCNF

- $R = (\text{branch-name}, \text{branch-city}, \text{assets}, \text{customer-name}, \text{loan-number}, \text{amount})$

<i>branch-name</i>	<i>branch-city</i>	<i>assets</i>	<i>customer-name</i>	<i>loan-number</i>	<i>amount</i>
Downtown	Brooklyn	9000000	Jones	L-17	1000
Redwood	Palo Alto	2100000	Smith	L-23	2000
Perryridge	Horseneck	1700000	Hayes	L-15	1500
Downtown	Brooklyn	9000000	Jackson	L-14	1500

- συναρτησιακές εξαρτήσεις:
 - $\text{branch-name} \rightarrow \text{assets}, \text{branch-city}$
 - $\text{loan-number} \rightarrow \text{amount}, \text{branch-name}$
- Αποσύνθεση
 - $R_1 = (\text{branch-name}, \text{branch-city}, \text{assets})$
 - $R_2 = (\text{branch-name}, \text{loan-number}, \text{amount})$
 - $R_3 = (\text{customer-name}, \text{loan-number})$

Περαιτέρω μελέτη



- Codd: A Relational Model of Data for Large Shared Data Banks (1970)
 - το κλασικό άρθρο απ' όπου ξεκίνησαν όλα !
- Stonebraker & Hellerstein: What Goes Around Comes Around.
 - μια ιστορική αναδρομή στα μοντέλα δεδομένων, από τη δεκαετία '60 (ιεραρχικό μοντέλο) μέχρι σήμερα (XML, κ.α.)