

Σχεσιακό Μοντέλο Δεδομένων

Πάυλος Εφραιμίδης

`pefraimi <at> ee.duth.gr`

Μοντέλα Δεδομένων

- Μοντέλα Δεδομένων
 - Σχεσιακό
 - Ιεραρχικό
 - Δικτυακό
- Το κυρίαρχο μοντέλο δεδομένων στις σύγχρονες βάσεις δεδομένων είναι το ΣΧΕΣΙΑΚΟ (και επεκτάσεις του)

Σχεσιακό Μοντέλο Δεδομένων

- Το σχεσιακό μοντέλο δεδομένων προτάθηκε από **Ted Codd** της *IBM Research* το 1970
- **Σχέση:** Η βάση του μοντέλου είναι η την έννοια της μαθηματικής σχέσης, που μοιάζει κατά κάποιον τρόπο με έναν πίνακα τιμών
- Η θεωρητική του βάση είναι η θεωρία συνόλων και ο κατηγορηματικός λογισμός πρώτης τάξης
- Γνώρισε άμεσα μεγάλη διάδοση λόγω
 - απλότητας
 - μαθηματικής θεμελίωσης

Τι θα πούμε

- Θα μιλήσουμε για
 - χαρακτηριστικά του μοντέλου,
 - τους περιορισμούς του,
- Για τη μαθηματική θεμελίωση του σχεσιακού μοντέλου με σχεσιακή άλγεβρα

ΕΝΝΟΙΕΣ του Σχεσιακού Μοντέλου (1)

- Μια βάση δεδομένων παριστάνεται ως ένα σύνολο από σχέσεις.
- Τι είναι μια σχέση:
 - Με απλά λόγια, μια σχέση μοιάζει με έναν πίνακα
 - Κάθε γραμμή του πίνακα παριστάνει μια συλλογή από τιμές δεδομένων που σχετίζονται με κάποιον τρόπο μεταξύ τους.

ΕΝΝΟΙΕΣ του Σχεσιακού Μοντέλου (2)

- Οι τιμές αυτές μπορεί να περιγράφουν μια οντότητα ή μια συσχέτιση του πραγματικού κόσμου. Δίνουμε όνομα στον πίνακα και στις στήλες του ώστε να είναι ευκολότερη η ερμηνεία του.
- Παράδειγμα: Πίνακας ΦΟΙΤΗΤΗΣ
 - Όνομα Σχέσης: ΦΟΙΤΗΤΗΣ
 - Γνωρίσματα: Όνομα, ΑΜ, Τηλ, Έτος, ..

Σχέση

- **Γνώρισμα μιας Σχέσης:** *Μια στήλη του πίνακα*
 - Όλες οι τιμές της στήλης είναι του ιδίου τύπου
 - Ο τύπος αυτό λέγεται **πεδίο ορισμού**
- **Πλειάδα μιας Σχέσης:** *Μια γραμμή του πίνακα*

Πεδίο Ορισμού

- **Πεδίο ορισμού D (Domain):** Ένα σύνολο από ατομικές τιμές
- **Παραδείγματα πεδίων ορισμού:**
 - Βαθμολογία ενός Φοιτητή: $0, \dots, 10$
 - Ηλικίες Εργαζόμενων: $18, \dots, 70$
 - Όνομα: Το σύνολο όλων των ονομάτων
 - Τοπικοί Αριθμοί Τηλεφώνου: 5-ψήφιοι
 - Ελληνικοί Αριθμοί τηλεφώνου: 10-ψήφιοι
- **Το πεδίο ορισμού ως περιορισμός:** Το πεδίο ορισμού είναι ένας από τους 4 βασικούς περιορισμούς του Σχεσιακού Μοντέλου Δεδομένων

Σχήμα μιας Σχέσης

- **ΣΧΗΜΑ ΣΧΕΣΗΣ:** Όνομα Σχέσης και Γνωρίσματα $R(A_1, \dots, A_n)$
- Το γνώρισμα A_i είναι το όνομα ενός ρόλου για ένα πεδίο ορισμού D στο σχήμα σχέσης R
- $D = \text{dom}(A_i)$ πεδίο ορισμού
- **Σχήμα Σχέσης**
 - Ένα σχήμα σχέσης περιγράφει μια σχέση
 - Το R είναι το όνομα της σχέσης
 - Βαθμός σχέσης είναι το πλήθος των γνωρισμάτων.
- Το σχήμα της σχέσης καθορίζει το είδος των δεδομένων της συγκεκριμένης σχέσης (πίνακα)
- **ΣΧΕΣΗ:** Τα δεδομένα
 - Ένα σύνολο από n -πλειάδες του σχήματος σχέσης
- **ΠΛΕΙΑΔΑ:** Μια διατεταγμένη λίστα από n τιμές, από τα αντίστοιχα πεδία ορισμού

ΜΑΘΗΜΑΤΙΚΟΣ ΟΡΙΣΜΟΣ

- *Σχέση $r(R)$ είναι μια μαθηματική σχέση βαθμού n στα πεδία ορισμού $dom(A1), dom(A2), \dots, dom(An)$ που είναι υποσύνολο του καρτεσιανού γινομένου των πεδίων ορισμού που ορίζουν την R :*
 - $r(R) \subseteq (dom(A1) \times dom(A2) \times \dots \times dom(An))$
 - Συνολικός αριθμός πλειάδων του καρτεσιανού γινομένου:
 $|dom(a1)| * |dom(A2)| * \dots * |dom(An)|$
- **Το πεδίο ορισμού D_i**
 - Μπορεί να είναι άπειρο
 - Μπορεί να επαναλαμβάνεται στο R
 - Περιέχει υποχρεωτικά μια ειδική τιμή **NULL**

Σχέση και Σχήμα Σχέσης

- Η **Σχέση** είναι ένα υποσύνολο το συνόλου όλων των πιθανών συνδυασμών
 - αντιστοιχεί στα δεδομένα του πίνακα
- Η σχέση μεταβάλλεται όμως το **σχήμα σχέσης R** γενικά δεν αλλάζει παρά μόνο σπάνια, πχ. πρόσθεση νέου γνωρίσματος
 - αντιστοιχεί στο **σχήμα (στα χαρακτηριστικά των πινάκων)**

Χαρακτηριστικά των Σχέσεων

- Η Σχέση είναι ένα σύνολο πλειάδων
Τα σύνολα δεν έχουν διάταξη, άρα
 - Δεν μας ενδιαφέρει η διάταξη των πλειάδων (πρακτικά θα έχουν αποθηκευτεί με κάποια σειρά)
 - Οι πλειάδες μιας σχέσης διαφέρουν μεταξύ τους (μοναδικότητα)
- Μπορούμε όταν παρουσιάζουμε μια σχέση να διατάσσουμε τις πλειάδες ως προς κάποιο γνώρισμα, πχ. Όνομα ή ΑΜ κτλ
- **ΠΛΕΙΑΔΑ:** Διατεταγμένη λίστα n-τιμών
 - Άρα έχει σημασία η διάταξη των τιμών σε μια πλειάδα.
 - Βέβαια σε λογικό επίπεδο μας απασχολεί απλά η σωστή αντιστοιχία ΤΙΜΗΣ-> στήλη-γνώρισμα

Ατομικές Τιμές (1)

- **Ατομικές Τιμές:** Κάθε τιμή σε μια πλειάδα είναι *ΑΤΟΜΙΚΗ*, δηλαδή δεν επιτρέπεται υποδιαίρεσή της σε συστατικά στοιχεία
- **ΠΡΟΒΛΗΜΑ για το Διάγραμμα ΟΣ:** Σύνθετα και Πλειότιμα γνωρίσματα δεν μπορούν να παρασταθούν αυτούσια
- **Αντιμετώπιση:** Μεγάλο μέρος της θεωρίας του σχεσιακού μοντέλου έχει κατά νου αυτήν την υπόθεση, που ονομάζεται *ΥΠΟΘΕΣΗ ΠΡΩΤΗΣ ΚΑΝΟΝΙΚΗΣ ΜΟΡΦΗΣ*

Λύσεις:

- **Σύνθετα Γνωρίσματα:** Με πολλά απλά γνωρίσματα.
- **Πλειότιμα Γνωρίσματα:** Με ξεχωριστή σχέση

Ατομικές Τιμές (2)

- **Διάγραμμα Ο-Σ:** Προσοχή, στο διάγραμμα Ο-Σ μπορούμε να τα χρησιμοποιούμε κανονικά!
 - Όταν έρθει η ώρα να απεικονισθεί το διάγραμμα ΟΣ σε σχεσιακό σχήμα, τότε κάνουμε ότι μετατροπές απαιτούνται
- **Τιμή NULL:** Όταν κάποιες τιμές δεν έχουν νόημα σε μια πλειάδα ή είναι άγνωστες: Τιμή *NULL*

Ερμηνεία μιας Σχέσης

- **Δήλωση ή βεβαίωση !**
Για παράδειγμα το Σχήμα Σχέσης ΦΟΙΤΗΤΗΣ δηλώνει ότι κάθε οντότητα ΦΟΙΤΗΤΗΣ έχει αυτά τα γνωρίσματα
- **ΠΛΕΙΑΔΑ:** *Η ύπαρξη μιας πλειάδας δηλώνει ότι ο συγκεκριμένος φοιτητής υπάρχει, είναι ΓΕΓΟΝΟΣ*
- **Οντότητες και Συσχετίσεις:** *Το σχεσιακό μοντέλο αναφέρεται ομοιόμορφα σε οντότητες και συσχετίσεις του Διαγράμματος Οντοτήτων-Συσχετίσεων*
- **ΚΑΤΗΓΟΡΗΜΑ (Μαθηματική Λογική):**
 - Ένα σχήμα σχέσης μπορεί να ερμηνευτεί ως κατηγορημα (PREDICATE):
 - Οι τιμές κάθε πλειάδας ερμηνεύονται ως οι τιμές που ικανοποιούν το κατηγορημα
 - Η ερμηνεία αυτή είναι χρήσιμη για την PROLOG (λογικό προγραμματισμό, κτλ.)

Συμβολισμοί του Σχεσιακού Μοντέλου

- Σχήμα σχέσης R βαθμού n: $R(A_1, A_2, \dots, A_n)$
- n-πλειάδα $t = \langle v_1, v_2, \dots, v_n \rangle$
- $t[A_i]$, $t.A_i = v_i$ στην πλειάδα t
- ΦΟΙΤΗΤΗΣ: Σχέση
- ΦΟΙΤΗΤΗΣ(Όνομα, ΑΜ, ..) Σχήμα Σχέσης
- Όνομα Γνωρισμάτων από τη σχέση: ΦΟΙΤΗΤΗΣ.Όνομα, ΦΟΙΤΗΤΗΣ.ΑΜ, ..
- Παράδειγμα:
 - $t = \langle \text{"Γιάννης Παπαδόπουλος"}, 255, 05410-12345 \rangle$
 - $t(\text{Όνομα}) =$
 - $t(\text{ΑΜ}) =$
 - $t(\text{Όνομα}, \text{ΑΜ}) =$

ΠΕΡΙΟΡΙΣΜΟΙ στο Σχεσιακό Μοντέλο

- 4 βασικοί περιορισμοί:
 - Περιορισμοί στο πεδίο ορισμού
 - Περιορισμοί κλειδιού
 - Περιορισμοί ακεραιότητας οντοτήτων
 - Περιορισμοί αναφορικής ακεραιότητας

Περιορισμός Πεδίου Ορισμού

- Η τιμή κάθε γνωρίσματος πρέπει να είναι μια ατομική τιμή από το πεδίο ορισμού.
- Συνηθισμένοι τύποι πεδίων ορισμού:
 - short integer, integer, long integer
 - float, double precision float
 - Χαρακτήρες
 - Συμβολοσειρές σταθερού και μεταβλητού μήκους
 - Ημερομηνία, Ώρα, ..
- Θα εξετάσουμε το θέμα του πεδίου ορισμού και κατά την παρουσίαση της SQL

Περιορισμοί κλειδιού

- Κάθε σχέση (πίνακας) είναι ένα σύνολο πλειάδων (γραμμών)
- Τα στοιχεία ενός (μαθηματικού) συνόλου είναι διακριτά
Επομένως: *Οι πλειάδες μιας σχέσης διαφέρουν υποχρεωτικά μεταξύ τους*

Περιορισμός Κλειδιού:

- Σε κάθε σχήμα σχέσης R συνήθως υπάρχουν υποσύνολα γνωρισμάτων τέτοια ώστε:
Κανένα ζευγάρι πλειάδων να μην μπορεί να έχει τις ίδιες τιμές σε όλα αυτά τα γνωρίσματα
- Έστω SK ένα τέτοιο σύνολο και t_1, t_2 δύο διαφορετικές πλειάδες, τότε: $t_1[SK] \neq t_2[SK]$

Υπερκλειδί

- **ΥΠΕΡΚΛΕΙΔΙ**
 - Ένα σύνολο SK από γνωρίσματα μοναδικό για μια σχέση, ονομάζεται **ΥΠΕΡΚΛΕΙΔΙ**
 - Κάθε σχέση έχει ένα προφανές υπερκλειδί: Το σύνολο όλων των γνωρισμάτων της
- Είναι πιθανό να αφαιρέσουμε ένα γνώρισμα από ένα υπερκλειδί και το νέο σύνολο να είναι και αυτό υπερκλειδί
- **ΚΛΕΙΔΙ**: Ένα υπερκλειδί από το οποίο δεν μπορεί να αφαιρεθεί κανένα γνώρισμα και να παραμείνει υπερκλειδί
 - **ΚΛΕΙΔΙ** είναι ένα ελάχιστο **ΥΠΕΡΚΛΕΙΔΙ**
- **Τιμή ενός κλειδιού**: Προσδιορίζει μονοσήμαντα την πλειάδα
- **Πως καθορίζεται το κλειδί**; Η ιδιότητα του κλειδιού είναι μια ιδιότητα του σχήματος σχέσης

Πρωτεύον Κλειδί

- **Κλειδιά:** Ένα σχήμα σχέσης μπορεί να έχει περισσότερα από ένα κλειδιά
- **Υποψήφια Κλειδιά:** Κάθε κλειδί ονομάζεται *ΥΠΟΨΗΦΙΟ ΚΛΕΙΔΙ* (*candidate key*)
- **Πρωτεύων Κλειδί:** Συνήθως ορίζουμε ένα από τα υποψήφια κλειδιά ως πρωτεύον (*primary key*)
- **Συμβολισμός:** Υπογραμμίζουμε τα γνωρίσματα που σχηματίζουν το πρωτεύον κλειδί
- **Επιλογή Πρωτεύοντος Κλειδιού:** Αν έχουμε πολλές επιλογές για το κλειδί, διαλέγουμε συνήθως μία με όσο δυνατό λίγα γνωρίσματα και προτιμούμε αριθμητικά γνωρίσματα

Άλλοι περιορισμοί

- **Σημασιολογική Ακεραιότητα:** Πχ. ο μισθός δεν μπορεί να υπερβαίνει κάποιο πλαφόν ή καλύτερα δεν μπορεί να είναι μικρότερος από το ελάχιστο νόμιμο
- **Άλλοι Μηχανισμοί:** Θα δούμε μελετώντας τη γλώσσα *SQL* μηχανισμούς όπως είναι οι *Triggers* και τα *Assertions* που μπορούν να χρησιμοποιηθούν για να υλοποιηθούν περιορισμοί
- **ΕΠΙΣΗΣ:** Οι παραπάνω περιορισμοί είναι περιορισμοί κατάστασης
- Υπάρχουν και μεταβατικοί περιορισμοί:
 - Πχ. Ο μισθός μπορεί μόνο να αυξάνει !!Μπορούν να υλοποιηθούν με δυναμικούς περιορισμούς, *triggers*, ..

Ακεραιότητα Οντοτήτων

- **Περιορισμός ακεραιότητας οντοτήτων:**
Δεν μπορεί η τιμή ενός γνωρίσματος που είναι κλειδί ή μέρος ενός κλειδιού (για σύνθετα κλειδιά) να έχει την τιμή NULL

Αναφορική Ακεραιότητα

- **Περιορισμός αναφορικής ακεραιότητας:** *Αφορά δύο σχέσεις και ορίζεται για να υπάρχει συνέπεια στις πλειάδες των δύο σχέσεων*
- **Με απλά λόγια:** *Μια πλειάδα μιας σχέσης που αναφέρεται σε μια άλλη σχέση, θα πρέπει να αναφέρεται σε μια πλειάδα της άλλης σχέσης που υπάρχει*
- **Παράδειγμα:** *Εάν ένας ΕΡΓΑΖΟΜΕΝΟΣ ανήκει στο τμήμα με κωδικό 100 θα πρέπει να υπάρχει στον πίνακα ΤΜΗΜΑ ένα τμήμα με τον κωδικό αυτό (100)*

Πηγές/Αναφορές

- Κεφάλαιο 5 από το Βιβλίο του Μαθήματος:
Θεμελιώσεις Αρχές Συστημάτων Βάσεων
Δεδομένων (Τόμος Α' - Έκδοση 4η), Elmasri
& Navathe (Μετάφραση: Εκδόσεις
ΔΙΑΥΛΟΣ)