Οργάνωση των δεδομένων σε πίνακες

ΣΤΟΧΟΣ

Στόχος του 1ου εισαγωγικού εργαστηρίου είναι η κατανόηση του τρόπου οργάνωσης της πληροφορίας σε δυσδιάστατους πίνακες. Η εν λόγω οργάνωση μπορεί να θεωρηθεί ανάλογη εκείνης όπου οι εγγραφές (records) καταχωρούνται σε αρχεία (files). Αυτήν τη φορά έχουμε πλειάδες (tuples)/ γραμμές οι οποίες καταχωρούνται σε πίνακες (tables). Ο κάθε ένας πίνακας, εκτός από τις γραμμές που περιέχει, χαρακτηρίζεται επίσης από τις στήλες του. Κάθε μία από τις τελευταίες μπορεί να θεωρηθεί ότι αποτελεί 'πεδίο' κάθε μίας γραμμής του πίνακα. Όλες οι γραμμές του πίνακα **έχουν την ίδια δομή**, αποτελούνται, δηλαδή, από το ίδιο πλήθος πεδίων κάθε ένα από τα οποία καταχωρεί **συγκεκριμένου τύπου δεδομένα**.

Το σύνολο των πινάκων, στον 'χώρο', ενός πληροφοριακού συστήματος λέμε ότι αποτελεί την αντίστοιχη, '***σχεσιακή βάση δεδομένων***'. Έχοντας κατανοήσει τα βασικά που αφορούν στην οργάνωση των δεδομένων στο τυπικό περιβάλλον σχεσιακής βάσης δεδομένων, θα προχωρήσουμε, σε επόμενα εργαστήρια, στη διεκπεραίωση 'πράξεων' επί του περιεχομένου των πινάκων. Η οργάνωση της πληροφορίας και η επεξεργασία της σε σχεσιακή βάση δεδομένων συνιστά τη λειτουργικότητα του σχεσιακού συστήματος βάσεων δεδομένων (Relational DataBase Management System)

ΣΧΕΤΙΚΟ ΕΚΠΑΙΔΕΥΤΙΚΟ ΥΛΙΚΟ

Η (ιστορική) εξέλιξη των συστημάτων διαχείρισης βάσεων δεδομένων σκιαγραφείται (σε γενικές γραμμές) στο βιβλίο του μαθήματος (ενότητα 1.2). Εκεί ο αναγνώστης μπορεί να κατανοήσει την ανάγκη ύπαρξης συστημάτων οργάνωσης και διαχείρισης της επεξεργασίας των δεδομένων ενός πληροφοριακού συστήματος σήμερα, ώστε να μπορούν να αναπτύσσονται και να συντηρούνται με παραγωγικό τρόπο πολύπλοκες εφαρμογές πληροφοριακών συστημάτων. Επίσης, στο Κεφάλαιο 3 (αγνοώντας τα περί κανονικοποίησης που θα εξετασθεί στη συνέχεια) εξηγούνται τα βασικά που έχουν να κάνουν με την έννοια της δομής του πίνακα, καθώς επίσης και τους περιορισμούς τους οποίους αυτή θέτει (π.χ. την ύπαρξη κύριου κλειδιού). Η άλγεβρα (πράξεις/ τελεστές) των πινάκων εξετάζεται στο Κεφάλαιο 4 του βιβλίου.

ΒΑΣΙΚΕΣ ΕΝΝΟΙΕΣ

Πίνακες/σχέσεις και βάσεις δεδομένων

Η οργάνωση της πληροφορίας σε σχεσιακό περιβάλλον βάσης δεδομένων μπορούμε να πούμε ότι υπόκειται στην ιεραρχία του Σχήματος 1.1:

Σχήμα 1.1 Από τα δεδομένα στη σχεσιακή βάση δεδομένων

Στο εν λόγω (σχεσιακό) περιβάλλον, είναι πολύ σημαντικό να γίνουν από την αρχή κατανοητές οι εξής έννοιες:

- *Όνομα Πίνακα*
- *Όνομα Στήλης*
- *Τύπος Στήλης*
- *Πλειάδα/ Γραμμή*
- *Κύριο Κλειδί*
- *Συσχετισμός πληροφορίας από περισσότερους από έναν πίνακες (Ξένο Κλειδί)*
- *Στοιχειώδεις Πράξεις (Τελεστές) στους Πίνακες: Προβολή & Επιλογή*
- *Μαζικού τύπου επεξεργασία στα περιεχόμενα ενός πίνακα*

ΠΕΡΙΒΑΛΛΟΝ ΓΙΑ ΕΞΑΣΚΗΣΗ

Έχουμε ένα μίνι πληροφοριακό σύστημα το οποίο έχει να κάνει με την καταχώρηση και διαχείριση στοιχείων κινηματογραφικών ταινιών (σε έκδοση βίντεο και σε DVD) και των ηθοποιών που παίζουν σε αυτές. Για λόγους που έχουν να κάνουν με τη μεθοδολογία σχεδιασμού μιας σχεσιακής βάσης δεδομένων (όπως ονομάζεται: **σχεσιακό σχήμα**), μεθοδολογία η οποία πρόκειται να εξεταστεί στη συνέχεια του μαθήματος, καταλήγουμε να οργανώσουμε τα δεδομένα σε δύο πίνακες:

Tainia (**titlos**, apoth_video, apoth_dvd, paragogi, t_video, t_dvd)

Symmetoxi (**titlos_tainias, eponimo_ithopoiou**, onoma_ithopoiou)

Στα παραπάνω, τα υπογραμμισμένα ονόματα στηλών υποδεικνύουν, στην κάθε μία περίπτωση, το κύριο κλειδί (σύνθετο στον πίνακα Symmetoxi). Σημειώνεται ότι οι στήλες apoth_video και apoth_dvd καταχωρούν στοιχεία που αφορούν στους διαθέσιμους αριθμούς (απόθεμα) ταινιών βίντεο και DVD στο κατάστημα, ενώ οι στήλες t_video και t_dvd καταχωρούν τις αντίστοιχες τιμές τους σε ευρώ.

Οι δύο πίνακες καταχωρούν δεδομένα, π.χ. ως εξής:

Tainia					
titlos	**Apoth_video**	**Apoth_dvd**	**paragogi**	**T_video**	**T_dvd**
A night in the opera	05	24	Fox	40	25
Pretty woman	10	18	Metro	45	32
Prince of Tides	05	10	Fox	35	30
The way we were	15	20	Columbia	41	37
Ενόραση	10	12	Σινεμάτικ	38	32
Η Φυγή	18	20	Σινεμάτικ	47	40
Ηρεμη δύναμη	28	19	3K	50	42
Ο κόσμος της Ελενας	08	08	Φωτοκίνηση	55	43
Βλέμμα του Οδυσσέα	10	18	3K	47	34

Symmetoxi		
Titlos_tainias	**Eponimo_ithopoiou**	**Onoma_ithopoiou**
A night in the opera	Marx Brothers	-
Pretty woman	Cormack	Stacey
Pretty woman	Gere	Richard
Pretty woman	Roberts	Julia
Prince of Tides	Redford	Robert
The way we were	Redford	Robert
The way we were	Robson	Thomas
The way we were	Rutherford	Chris
The way we were	Streisand	Barbra
The way we were	Whitney	Maggie
Ενόραση	Καραφωτιάς	Χρήστος
Ενόραση	Σφυρίου	Ηρώ
Η Φυγή	Καζάς	Ιάσων
Η Φυγή	Σφυρίου	Ηρώ
Ηρεμη δύναμη	Καζάς	Ιάσων
Ηρεμη δύναμη	Κλέους	Περσεφόνη
Ηρεμη δύναμη	Λάμπρου	Νιόβη
Ηρεμη δύναμη	Σφυρίου	Ηρώ
Το βλέμμα του Οδυσσέα	Καραφωτιάς	Χρήστος
Το βλέμμα του Οδυσσέα	Κουτρής	Στέφανος
Το βλέμμα του Οδυσσέα	Λάζου	Αργυρώ

Εικόνα 1.1 Οι πίνακες Tainia και Symmetoxi και το περιεχόμενό τους

Για τη δημιουργία των δύο πινάκων θα χρησιμοποιήσουμε το αρχείο create_schema_two_tables.txt του **κώδικα SQL** που συνοδεύει αυτό το εργαστήριο. Στη σύνταξη της αντίστοιχης εντολής (DDL) SQL CREATE TABLE, τα πεδία titlos και paragogi του πίνακα Tainia και τα πεδία titlos_tainias, eponimo_ithopoiou και onoma_ithopoiou δηλώνονται να είναι τύπου varchar(50). Ο τύπος *varchar* χρησιμοποιείται στις βάσεις δεδομένων για την αποθήκευση αλφαριθμητικών τιμών. Στη γενική περίπτωση οι βάσεις δεδομένων υποστηρίζουν εκτός του τύπου varchar και τον τύπο *char* για την αποθήκευση αλφαριθμητικών τιμών.

Σύγκριση των τύπων varchar και char

Ομοιότητα: Κατά τη δήλωση του τύπου των δεδομένων μιας στήλης πίνακα, αναγράφεται σε παρένθεση το μέγιστο πλήθος χαρακτήρων που θα μπορεί να έχει η κάθε μία αντίστοιχη συμβολοσειρά/τιμή. Π.χ. αν επιθυμούμε να επιτρέπονται έως και 50 χαρακτήρων συμβολοσειρές, δηλώνουμε varchar(50) ή char(50).

Διαφορά: Ο τύπος char δεσμεύει ακριβώς το καθορισμένο πλήθος χαρακτήρων ανεξάρτητα από το μήκος (σε χαρακτήρες) της συμβολοσειράς/τιμής. Π.χ. για το παραπάνω παράδειγμα των 50 χαρακτήρων θα δεσμεύονται ακριβώς 50 χαρακτήρες για την κάθε μία συμβολοσειρά/τιμή. Αντίθετα, ο τύπος varchar δεσμεύει το πλήθος των χαρακτήρων της συμβολοσειράς και όχι παραπάνω, με την προϋπόθεση ότι δεν παραβιάζεται το άνω όριο που έχει οριστεί στη δήλωση του τύπου. Για παράδειγμα, όταν η συμβολοσειρά "Λαμία" καταχωρείται σε πεδίο στήλης που έχει δηλωθεί να καταχωρεί δεδομένα τύπου varchar(20) δεσμεύονται πέντε (5) χαρακτήρες στη μνήμη του συστήματος, ενώ όταν η ίδια συμβολοσειρά καταχωρείται ως δεδομένα τύπου char(20) τότε

δεσμεύονται είκοσι (20) χαρακτήρες. Εννοείται ότι στην τελευταία περίπτωση το σύστημα καταχωρεί και διαχειρίζεται μία συμβολοσειρά μήκους είκοσι χαρακτήρων όπου οι πρώτοι πέντε είναι το "Λαμία" και οι υπόλοιποι δεκεπέντε είναι επαναλήψεις του κενού (blank) χαρακτήρα.

Ενδιαφέρον παρουσιάζει η έννοια του **κυρίως κλειδιού (*primary key)*** του κάθε ενός πίνακα. Κύριο κλειδί πίνακα συνιστούν η στήλη ή ο συνδυασμός στηλών πίνακα, η (συνδυασμένη) τιμή του οποίου παίρνει μοναδικές τιμές σε κάθε μία γραμμή του πίνακα. Η στήλη ή οι στήλες που συνιστά(-ούν) το κύριο κλειδί του πίνακα υπόκειται στον περιορισμό του να μην μπορεί να παραλείπεται και να μένει χωρίς τιμή όταν εκτελούνται ενέργειες εισαγωγής νέου ή ενημέρωσης του ήδη υπάρχοντος περιεχομένου του πίνακα. Η μη ύπαρξη τιμής σε πεδίο/κυψέλη πίνακα ισοδυναμεί με **κατάσταση *null* (= τίποτα)**. Η στήλη ή οι στήλες που απαρτίζουν το κύριο κλειδί πίνακα δηλώνονται να είναι ***not null***, δηλαδή: τους απαγορεύεται να μένουν χωρίς να καταχωρούν κάποια τιμή (ισοδύναμα: τους απαγορεύετεια να καταχωρούν την "τιμή" null).

Σχολιασμός των πινάκων Tainia και Simmetoxi

Παρατηρώντας το περιεχόμενο των πινάκων Tainia και Symmetoxi, παρατηρούμε/ σχολιάζουμε τα εξής:

1. Κάθε ένας τους ικανοποιεί τον περιορισμό της μοναδικής τιμής του κύριου κλειδιού (***ακεραιότητα οντοτήτων*** - entity integrity)
2. Ο καθορισμός του κύριου κλειδιού στον πίνακα συνιστά περιορισμό με τον οποίο οφείλει να "ζει" ο πίνακας στη συνέχεια: δεν πρέπει να υπάρξει ποτέ περίπτωση όπου δύο γραμμές του θα καταχωρούν ίδια τιμή στο κλειδί.
3. Το πληροφοριακό περιεχόμενο των δύο πινάκων μπορεί να συνδυαστεί με 'οδηγό' το ζεύγος των στηλών (Tainia.*Titlos*, Symmetoxi.*titlos_tainias*). Οι δύο εν λόγω στήλες καταχωρούν ιδίου τύπου δεδομένα (όχι απαραίτητα και τις ίδιες, ακριβώς, τιμές). Η στήλη Symmetoxi.*titlos_tainias,* το περιεχόμενο της οποίας παραπέμπει σε περιεχόμενο-τιμή του κύριου κλειδιού του πίνακα Tainia (άρα: σε **μία και μόνη** γραμμή του τελευταίου) ονομάζεται ***ξένο κλειδί*** (για τον πίνακα Symmetoxi)
4. Δεν υπάρχει τίποτα που να μας εμποδίζει να συνδυάσουμε πληροφορία από τους δύο πίνακες με 'οδηγό' άλλα ζεύγη στηλών όπου καταχωρούνται ίδιου τύπου τιμές. Ωστόσο: υπάρχει νόημα στο αποτέλεσμα το οποίο προκύπτει αν χρησιμοποιηθούν άλλα ζεύγη;
5. Τα ζεύγη (όχι όλα!) των ομοειδών στηλών (γενίκευση: και συνδυασμών τους) επάνω στα οποία μπορούν να 'κτίζονται' συνδυασμοί του περιεχομένου των (όποιων) δύο πινάκων σε μία σχεσιακή βάση λέμε ότι υλοποιούν ***συσχετίσεις*** μεταξύ των πινάκων. **Προσοχή:** εννοείται ότι μιλάμε για συνδυασμούς όπου το παραγόμενο αποτέλεσμα έχει αξία (δηλ. νόημα)...
6. Όπου υπάρχει συσχέτιση, υπάρχει ζευγάρι του τύπου (κύριο κλειδί, ξένο κλειδί).
7. Δεν είναι προφανές ακόμη, όμως υπάρχει ένας ισχυρός κανόνας στις σχεσιακές βάσεις δεδομένων: οι τιμές τις οποίες παίρνει ένα ξένο κλειδί πρέπει να έχουν αντίστοιχες (δηλ. να έχουν 'αντίκρισμα') οι οποίες να καταχωρούνται ήδη στον πίνακα όπου υπάρχει το (αντίστοιχο) κύριο κλειδί. Για παράδειγμα, εξετάζοντας τα δεδομένα των πινάκων Tainia και Symmetoxi: η στήλη Symmetoxi.*titlos_tainias* δεν επιτρέπεται να καταχωρεί τιμές οι οποίες να μην είναι ήδη καταχωρημένες στη στήλη Tainia.*titlos*. **Προσοχή:** το αντίστροφο δεν απαγορεύεται. Ο εν λόγω περιορισμός επιβάλλεται από (ισοδύναμα: συνιστά) την ***αναφορική ακεραιότητα των δεδομένων*** (referential integrity).

Περιβάλλον εργασίας (λογισμικό)

Για την εκτέλεση των εντολών θα χρησιμοποιηθεί τα περιβάλλοντα **IBM DB2 Express-C** και **IBM Data Studio**. Και τα δύο διατίθενται **ελεύθερα** από την IBM και μπορούν να ανακτηθούν από τη διεύθυνση **http://www-01.ibm.com/software/data/db2/express-c/**. Μάλιστα, η (τρέχουσα) έκδοση 10.5 του IBM DB2 Express-C ενσωματώνει, μεταξύ άλλων, το IBM Data Studio. Ελεύθερα διατίθεται επίσης τα αντίστοιχα βιβλία **Getting Started with DB2 Express-C** και **Getting Started with IBM Data Studio for DB2**.

Σήμερα (Οκτώβριος 2013) η IBM DB2 Express-C διατίθεται στην υπ' αριθμόν 10.5 έκδοσή της. Μπορεί να εγκατασταθεί σε όλες τις δημοφιλείς πλατφόρμες λειτουργικών συστημάτων (MS-Windows, Linux, Mac OS-X, κλπ.). Το βιβλίο Getting Started with DB2 Express-C συμπεριλαμβάνει οδηγίες για την εγκατάσταση και τη χρήση του λογισμικού, πλην όμως αναφέρεται στην αμέσως προγενέστερη έκδοσή του (9.7). Μία Ελληνική έκδοση οδηγιών χρήσης του περιβάλλοντος IBM Data Studio αποτελεί το εγχειρίδιο "**Οδηγός εγκατάστασης και χρήσης του IBM Data Studio**".

Ευέλικτη εναλλακτική λύση χρήσης των IBM DB2 Express-C και IBM Data Studio συνιστά η εικονική μηχανή (virtual machine) DebianDB της διεθνούς σύμπραξης **DBTechNet** όπου συμμετέχει ενεργά το τμήμα Μηχανικών Πληροφορικής του ΑΤΕΙ-Θ. Για την εγκατάσταση και χρήση της εικονικής μηχανής απαιτούνται (α) η εγκατάσταση του λογισμικού **Oracle VirtualBox και του αντίστοιχου extension pack**, και (β) η ύπαρξη μνήμης 2 GBytes RAM (κατ' ελάχιστον!) στον υπολογιστή που χρησιμοποιείται. Κατά τα άλλα, το λογισμικό Oracle VirtualBox είναι διαθέσιμο για όλα τα δημοφιλή λειτουργικά συστήματα υπολογιστών. Στην ουσία, χρήση της εικονικής μηχανής DBTechNet σημαίνει χρήση ενός έτοιμου περιβάλλοντος Debian Linux "κάτω" από το λειτουργικό σύστημα του υπολογιστή ο οποίος το φιλοξενεί. Το λειτουργικό περιβάλλον Debian Linux του DBTechNet συμπεριλαμβάνει (προεγκατεστημένο!) ένα ευρύ φάσμα λογισμικού βάσεων δεδομένων: IBM DB2, MySQL, Oracle, PostgreSQL, και Pyrrho DBMS. Η ευελιξία του συνίσταται στην απρόσκοπτη μεταφερσιμότητά του από υπολογιστή σε υπολογιστή και, φυσικά, στην απλή διαδικασία εγκατάστασής του: ο φοιτητής χρειάζεται να εγκαταστήσει στον υπολογιστή του μόνον το λογισμικό Oracle VirtualBox και τίποτα άλλο. ¨ολα τα υπόλοιπα είναι προεγκατεστημένα στην εικονική μηχανή DBTechNet. Για την τελευταία έχει συνταχθεί και είναι διαθέσιμος ο σχετικός "**Γρήγορος οδηγός για το εργαστήριο DebianDB**" στο κείμενο του οποίου συμπεριλαμβάνονται οι σχετικοί υπερ-σύνδεσμοι (links) για την ανάκτηση και την εγκατάσταση της εικονικής μηχανής.

ΑΣΚΗΣΕΙΣ

1. Δημιουργήστε στο περιβάλλον του Data Studio της DB2 τους πίνακες Tainia και Symmetoxi. Για τη δημιουργία των δύο πινάκων θα χρησιμοποιήσετε το αρχείο create_schema_two_tables.txt του **κώδικα SQL** που συνοδεύει αυτό το εργαστήριο.
2. Δημιουργήστε ένα νέο πίνακα με όνομα Tainies_Ithopoioi, η δομή του οποίου συμπεριλαμβάνει όλες τις στήλες των πινάκων Tainia και Symmetoxi. Χρησιμοποιήστε τα αρχεία create_schema_single_table.txt από τον **κώδικα SQL** που συνοδεύει αυτό το εργαστήριο για να καταχωρήσετε σε αυτόν το ΣΥΝΟΛΟ της πληροφορίας της Εικόνας 1.1
3. Σχολιάστε: που υπερέχει (και που υπολείπεται) το σχήμα των δύο πινάκων (Tainia και Symmetoxi) εκείνου του ενός μόνον πίνακα (Tainies_Ithopoioi); Για να κάνετε τη σύγκριση, λάβετε υπόψη **και** το μέγεθος του κάθε ενός σχήματος (σε δεδομένα που καταχωρούνται) **και** αυτό το οποίο αντιλαμβάνεστε να αποτελεί 'κόστος επεξεργασίας' των δεδομένων.
4. Τελικά, λαμβάνοντας υπόψη τα του ερωτήματος 3 παραπάνω και όσα αφορούν στο συνδυασμό του περιεχομένου δύο ή περισσοτέρων πινάκων, θα μπορούσαμε να έχουμε το σύνολο της πληροφορίας που αφορά το παράδειγμά μας σε περισσότερους από δύο πίνακες, όχι μόνο στους Tainia και Symmetoxi; Σημείωση: Θεωρήστε ότι δεν υπάρχουν δύο ηθοποιοί οι οποίοι να έχουν το ίδιο επώνυμο (παραδοχή, απλούστευση).
5. Επιστρέφοντας στο σχήμα των δύο πινάκων (Εικόνα 1.1): προσθέστε μία νέα γραμμή στον πίνακα Symmetoxi σε τρόπο ώστε να παραβιάζεται η αναφορική ακεραιότητα των δεδομένων
6. Στο σχήμα των δύο πινάκων (Εικόνα 1.1): θα μπορούσε η προσθήκη μίας νέας γραμμής στον πίνακα Tainia να οδηγήσει σε παραβίαση της αναφορικής ακεραιότητας των δεδομένων; Ναι, όχι, γιατί (να αιτιολογηθεί η απάντησή σας).