

Κεφάλαιο 5

Σχεσιακή Άλγεβρα

Στο κεφάλαιο αυτό παρουσιάζονται οι θεμελιώδεις πράξεις της σχεσιακής άλγεβρας που χρησιμοποιούνται για τη διαχείριση των δεδομένων των πινάκων μιας σχεσιακής βάσης δεδομένων.

Η σχεσιακή άλγεβρα (relational algebra) ορίζεται ως ένα σύνολο πράξεων με τις οποίες μπορούμε να διαχειριστούμε τα δεδομένα των πινάκων μιας σχεσιακής βάσης δεδομένων. Αυτή η διαχείριση περιλαμβάνει την επιλογή γραμμών και στηλών οι οποίες πληρούν κάποια συγκεκριμένα κριτήρια, τόσο από ένα απλό πίνακα όσο και από το συνδυασμό δύο ή περισσότερων πινάκων. Το αποτέλεσμα της εφαρμογής αυτών των πράξεων είναι ένας νέος πίνακας πάνω στον οποίο μπορούμε να εφαρμόσουμε και άλλες τέτοιες πράξεις. Με τον τρόπο αυτό είναι τελικά δυνατή η πραγματοποίηση οποιασδήποτε ενέργειας, όσο πολύπλοκη και αν είναι.

Από τη βασική θεωρία είναι γνωστό πως ένας πίνακας ορίζεται ως ένα μη διατεταγμένο σύνολο πλειάδων (tuples). Εφ' όσον λοιπόν ο πίνακας διαθέτει όλες τις ιδιότητες ενός συνόλου, είναι προφανές πως μπορούμε να εφαρμόσουμε πάνω στους πίνακες της βάσης δεδομένων, όλες τις γνωστές πράξεις από τη θεωρία συνόλων – οι πράξεις αυτές είναι η ένωση (union), η τομή (intersection), η διαφορά (difference) και το καρτεσιανό γινόμενο (Cartesian product). Η άλλη κατηγορία πράξεων που αποτελούν τμήμα της σχεσιακής άλγεβρας, περιλαμβάνει διαδικασίες οι οποίες έχουν οριστεί ειδικά για το σχεσιακό μοντέλο βάσεων δεδομένων, όπως είναι η επιλογή (select), η προβολή (project), η σύζευξη (join) και η διαίρεση (division). Η αναλυτική περιγραφή όλων αυτών των πράξεων της σχεσιακής άλγεβρας, αποτελεί αντικείμενο των σελίδων που ακολουθούν.

Η ΠΡΑΞΗ ΤΗΣ ΕΠΙΛΟΓΗΣ (SELECT OPERATION)

Η πράξη της επιλογής χρησιμοποιείται για να επιλέξει ένα σύνολο από πλειάδες κάποιου πίνακα οι οποίες ικανοποιούν μια συγκεκριμένη συνθήκη (selection condition). Επομένως η πραγματοποίησή της, απαιτεί τον καθορισμό τόσο του ονόματος του πίνακα επί του οποίου θα εφαρμοσθεί – ας σημειωθεί πως η πράξη αυτή εφαρμόζεται μόνο σε ένα πίνακα (unary operation)– όσο και της συνθήκης επιλογής που καθορίζει τις πλειάδες που θα εμφανιστούν στο τελικό αποτέλεσμα. Έτσι η πράξη αυτή καλείται με τη σύνταξη

$\sigma_{\langle \text{selection condition} \rangle} (\langle \text{relation name} \rangle)$

Για παράδειγμα, εάν επιθυμούμε να ανακτήσουμε τους εργαζόμενους της εταιρείας που εργάζονται στο τμήμα με κωδικό 4 θα καλέσουμε την πράξη με τη μορφή

$$\sigma_{\text{DNO}=4}(\text{EMPLOYEE})$$

ενώ για να λάβουμε όλους τους εργαζόμενους της εταιρείας με μισθό μεγαλύτερο από 30000, θα γράψουμε

$$\sigma_{\text{SALARY}>30000}(\text{EMPLOYEE})$$

Στην παραπάνω σύνταξη, η συνθήκη επιλογής μπορεί να είναι όσο πολύπλοκη επιθυμούμε, ενώ επιπλέον έχουμε και τη δυνατότητα να καταχωρήσουμε το αποτέλεσμα σε ένα ενδιάμεσο πίνακα δια της χρήσης του τελεστή «←». Έτσι η πράξη

$$\text{RESULT} \leftarrow \sigma_{\text{DNO}=5 \text{ AND } \text{SALARY}>25000}(\text{EMPLOYEE})$$

επιλέγει τις πλειάδες του πίνακα **EMPLOYEE** που εργάζονται στο τμήμα με κωδικό 5 και ο μισθός τους είναι μεγαλύτερος από 25000, και στη συνέχεια, καταχωρεί το αποτέλεσμα στον πίνακα **RESULT**.

Εκτός από το σύμβολο της ισότητας, στην παραπάνω συνθήκη επιλογής, μπορούν να χρησιμοποιηθούν όλοι οι γνωστοί **τελεστές σύγκρισης** {<, ≤, >, ≥, ≠} με την προϋπόθεση πως **οι τιμές που παίρνουν τα πεδία πάνω στα οποία εφαρμόζονται, να είναι διατεταγμένες**, έτσι ώστε να είναι δυνατή η εφαρμογή αυτών των τελεστών – αυτό ισχύει πάντα όταν τα πεδία αυτά είναι **αριθμοί ή ημερομηνίες**. Στην περίπτωση κατά την οποία τα πεδία της συνθήκης επιλογής είναι **συμβολοσειρές**, τότε χρησιμοποιείται η **αλφαβητική ταξινόμηση** για τη σύγκρισή τους, ενώ όταν είναι **αλφαριθμητικά**, η σύγκρισή τους γίνεται με βάση τους κώδικες **ASCII** των χαρακτήρων τους. Τέλος, όταν αναφερόμαστε σε πεδία τιμών τα οποία δεν είναι δυνατόν να ταξινομηθούν, μπορούμε να χρησιμοποιήσουμε μόνο τους τελεστές {=, ≠}. Για παράδειγμα εάν κάποιο από τα πεδία αναπαριστά κάποιο χρώμα που παίρνει τιμές από το σύνολο **Color = {red, green, blue, black, yellow, white}**, μπορούμε να χρησιμοποιήσουμε μόνο τους δύο προαναφερόμενους τελεστές, διότι δεν έχει νόημα να πούμε πως κάποιο χρώμα είναι μεγαλύτερο ή μικρότερο από κάποιο άλλο – αλλά μπορούμε να γράψουμε **color = black ή color ≠ red**.

Εκτός από τους τελεστές σύγκρισης, σε μια συνθήκη επιλογής μπορούμε να χρησιμοποιήσουμε και τους **λογικούς τελεστές AND, OR και NOT**. Στην ειδική περίπτωση του τελεστή **AND** που χρησιμοποιείται για να συσχετίσει δύο πράξεις σύγκρισης τιμών, το αποτέλεσμα που θα προκύψει είναι το ίδιο με εκείνο που θα παίρναμε εάν εφαρμόζαμε τη **SELECT** με την πρώτη συνθήκη, και στον πίνακα που θα προέκυπτε, εφαρμόζαμε ξανά τη **SELECT**, αυτή τη φορά με τη δεύτερη συνθήκη. Για παράδειγμα η πράξη

$$\text{RESULT} \leftarrow \sigma_{\text{DNO}=5 \text{ AND } \text{SALARY}>25000}(\text{EMPLOYEE})$$

είναι ισοδύναμη με την ακολουθία πράξεων

$$\text{RESULT} \leftarrow \sigma_{\text{DNO}=5} \left(\sigma_{\text{SALARY}>25000} (\text{EMPLOYEE}) \right)$$

Ο τρόπος με τον οποίο λειτουργεί η **SELECT**, είναι ο εξής: η εντολή διαβάσει μια - μια τις πλειάδες του πίνακα και σε κάθε μία από αυτές εφαρμόζει τη συνθήκη επιλογής. Για να το κάνει αυτό, διαβάσει τις τιμές των πεδίων που εμφανίζονται στη συνθήκη επιλογής, και τις συγκρίνει με τις αντίστοιχες που καθορίζονται στην εν λόγω συνθήκη. Εάν αυτή η συνθήκη ικανοποιείται, τότε αυτή η πλειάδα προστίθεται στο τελικό αποτέλεσμα – στην αντίθετη περίπτωση απορρίπτεται. Επίσης είναι προφανές πως αφού η **SELECT** επιστρέφει πλειάδες, ο πίνακας που θα προκύψει ως αποτέλεσμα, θα έχει την ίδια δομή με τον πίνακα πάνω στον οποίο εφαρμόζεται η εντολή.

Η ΠΡΑΞΗ ΤΗΣ ΠΡΟΒΟΛΗΣ (PROJECT OPERATION)

Η πράξη της προβολής χρησιμοποιείται για να επιλέξει όχι γραμμές του πίνακα όπως η **SELECT**, αλλά στήλες. Με άλλα λόγια ο πίνακας που θα προκύψει ως αποτέλεσμα εφαρμογής της **PROJECT**, δεν θα έχει την ίδια δομή με τον πίνακα πάνω στον οποίο εφαρμόζεται, αλλά θα έχει γενικά λιγότερες στήλες.

Εφ' όσον η **PROJECT** χρησιμοποιείται για την επιλογή στηλών από ένα πίνακα, θα πρέπει να κληθεί με τρόπο που να καθορίζει τόσο τον πίνακα πάνω στον οποίο θα εφαρμοσθεί, όσο και τις στήλες αυτού του πίνακα που θα επιστρέψει. Έτσι, η εντολή **PROJECT** καλείται με τη μορφή

$$\pi_{\langle \text{attribute list} \rangle} (\langle \text{relation name} \rangle)$$

Ένα παράδειγμα χρήσης της **PROJECT** είναι η κλήση της με τη μορφή

$$\pi_{\text{LNAME, FNAME, SSN, ADDRESS}} (\text{EMPLOYEE})$$

που θα επιστρέψει αυτές τις τέσσερις στήλες του πίνακα **EMPLOYEE** οδηγώντας έτσι στο αποτέλεσμα που ακολουθεί στη συνέχεια. Επίσης είναι προφανές πως η **PROJECT** - όπως και η **SELECT** – εφαρμόζεται μόνο σε ένα πίνακα κάθε φορά.

FNAME	LNAME	SSN	ADDRESS
John	Smith	123456789	731 Fondren, Houston, TX
Franklin	Wong	333445555	638 Vass, Houston, TX
Alicia	Zelaya	999887777	3321 Castle, Spring, TX
Jennifer	Wallace	987654321	291 Berry, Bellaire, TX
Ramesh	Narayan	666884444	975 Fire Oak, Humble, TX
Joyce	English	453453453	5631 Rice, Houston, TX
Ahmad	Jabbar	987987987	980 Dallas, Houston, TX
James	Borg	888665555	450 Stone, Houston, TX

Στην περίπτωση κατά την οποία η λίστα των στηλών που θέλουμε να ανακτήσουμε δεν περιλαμβάνει το πρωτεύον κλειδί του πίνακα, υπάρχει η πιθανότητα το τελικό αποτέλεσμα να περιέχει επαναλαμβανόμενες εγγραφές. Στην περίπτωση αυτή η

PROJECT κρατά μόνο ένα αντίγραφο από κάθε εγγραφή, οδηγώντας έτσι σε ένα έγκυρο αποτέλεσμα. Η διαδικασία αυτή είναι γνωστή και ως **εξάλειψη διπλοεγγραφών (duplicate elimination)**.

Επειδή η εντολή **SELECT** επιστρέφει **πλειάδες** του πίνακα ενώ η **PROJECT** επιστρέφει **στήλες**, είναι προφανές πως εάν εφαρμόσουμε και τις δύο εντολές ταυτόχρονα πάνω σε κάποιο πίνακα, θα πάρουμε μόνο ένα τμήμα του το οποίο θα περιέχει ένα συγκεκριμένο αριθμό γραμμών και στηλών. Ένα παράδειγμα συνδυασμένης χρήσης της **SELECT** και της **PROJECT** είναι ο πίνακας που προκύπτει από την πράξη

$$\pi_{\text{LNAME, FNAME, SALARY}}(\sigma_{\text{DNO=5 AND SALARY>25000}}(\text{EMPLOYEE}))$$

και περιέχει **τα ονόματα και το μισθό των υπαλλήλων της εταιρείας που εργάζονται στο τμήμα με κωδικό 5 και έχουν μισθό μεγαλύτερο από 25000**. Το αποτέλεσμα αυτής της συνδυασμένης χρήσης των δύο εντολών παρουσιάζεται στον ακόλουθο πίνακα:

FNAME	LNAME	SALARY
John	Smith	30000
Franklin	Wong	40000
Ramesh	Narayan	38000

Ας σημειωθεί πως εάν το επιθυμούμε, έχουμε τη δυνατότητα να μην χρησιμοποιήσουμε τις δύο εντολές με αυτό το συνδυασμένο τρόπο σύνταξης, αλλά να εκτελέσουμε την πράξη σε δύο βήματα, και χρησιμοποιώντας ένα ενδιάμεσο πίνακα. Έτσι μπορούμε να γράψουμε

$$\text{DEP5_EMPS} \leftarrow \sigma_{\text{DNO=5 AND SALARY>25000}}(\text{EMPLOYEE})$$

$$\text{RESULT} \leftarrow \pi_{\text{LNAME, FNAME, SALARY}}(\text{DEP5_EMPS})$$

με τον πίνακα **DEP5_EMPS** να περιέχει το αποτέλεσμα εφαρμογής της **SELECT** πάνω στον πίνακα **EMPLOYEE**, και τον πίνακα **RESULT** να περιέχει το αποτέλεσμα της **PROJECT** πάνω στον πίνακα **DEP5_EMPS**.

Τέλος είναι σημαντικό να αναφέρουμε πως ο νέος πίνακας που προκύπτει από την χρήση των παραπάνω εντολών **διατηρεί τα ίδια ονόματα πεδίων με τους πίνακες από τους οποίους προέρχεται**. Εάν ωστόσο επιθυμούμε να αλλάξουμε τα ονόματα αυτών των πεδίων, μπορούμε να το κάνουμε γράφοντάς τα εντός παρενθέσεων μετά το όνομα του πίνακα. Έτσι στην προηγούμενη περίπτωση, τα πεδία του πίνακα **RESULT**, φέρουν τα ονόματα **LNAME, FNAME** και **SALARY**, ενώ εάν γράψουμε

$$\text{RESULT}(\text{FIRSTNAME, LASTNAME, SALARY}) \leftarrow$$

$$\pi_{\text{LNAME, FNAME, SALARY}}(\text{DEP5_EMPS})$$

η δομή του πίνακα **RESULT** θα είναι η {**FIRSTNAME, LASTNAME, SALARY**}. Ας σημειωθεί πως οι δύο τελευταίες παρατηρήσεις σχετικά με τη βηματική εκτέλεση των εντολών και τη μετονομασία των πεδίων του τελικού πίνακα, δεν ισχύουν μόνο για τις πράξεις της επιλογής και της προβολής, αλλά για όλες τις πράξεις της σχεσιακής άλγεβρας.

ΠΡΑΞΕΙΣ ΑΝΑΜΕΣΑ ΣΤΟΥΣ ΠΙΝΑΚΕΣ ΑΠΟ ΤΗ ΘΕΩΡΙΑ ΣΥΝΟΛΩΝ

Στην ενότητα αυτή θα μελετήσουμε τις βασικές πράξεις που μπορούμε να πραγματοποιήσουμε ανάμεσα σε δύο σύνολα, και οι οποίες όπως έχουμε ήδη αναφέρει, μπορούν να εφαρμοσθούν και στους πίνακες μιας βάσης δεδομένων, διότι σύμφωνα με τον ορισμό τους **οι πίνακες είναι σύνολα από πλειάδες**. Οι πράξεις αυτές είναι η **ένωση (union)**, η **τομή (intersection)**, η **διαφορά (difference)** και το **καρτεσιανό γινόμενο (Cartesian product)**. Οι πράξεις αυτές είναι **δυναδικές**, διότι εφαρμόζονται πάντα ανάμεσα σε δύο πίνακες της βάσης.

Απαραίτητη προϋπόθεση για την εφαρμογή των τριών πρώτων πράξεων είναι οι πίνακες επί των οποίων εφαρμόζονται, να είναι **συμβατοί ως προς την ένωση (union compatible)**. Δύο πίνακες $R(A_1, A_2, \dots, A_n)$ και $S(B_1, B_2, \dots, B_n)$ θεωρούνται **συμβατοί ως προς την ένωση**, εάν έχουν το ίδιο πλήθος πεδίων, n , και επιπλέον ισχύει η συνθήκη $\text{dom}(A_i) = \text{dom}(B_i)$ ($1 \leq i \leq n$), που σημαίνει πως **τα αντίστοιχα πεδία των δύο πινάκων να έχουν το ίδιο πεδίο τιμών**. Στην περίπτωση αυτή οι πράξεις της **ένωσης**, της **τομής** και της **διαφοράς** των πινάκων **R** και **S**, ορίζονται με τον ακόλουθο τρόπο:

Η **ένωση** δύο πινάκων, $R \cup S$, είναι ένας νέος πίνακας που περιλαμβάνει **όλες τις πλειάδες των δύο πινάκων, οι οποίες μπορεί να ανήκουν μόνο στον R, μόνο στον S ή και στους δύο πίνακες ταυτόχρονα**. Εάν μετά την ένωση των δύο πινάκων υπάρχουν εγγραφές που εμφανίζονται περισσότερες από μία φορές, αυτές απομακρύνονται από το τελικό αποτέλεσμα, έτσι ώστε να υπάρχει μόνο ένα αντίγραφο από κάθε εγγραφή.

Η **τομή** δύο πινάκων, $R \cap S$, είναι ένας νέος πίνακας, που περιλαμβάνει **όλες τις πλειάδες οι οποίες είναι κοινές στους δύο πίνακες**.

Τέλος η **διαφορά** των πινάκων, $R - S$, είναι ένας νέος πίνακας, που περιλαμβάνει **όλες τις πλειάδες που ανήκουν στον R, αλλά δεν ανήκουν στον S**.

Για τις πράξεις της τομής και της ένωσης, ισχύει η **αντιμεταθετική ιδιότητα**, μπορούμε δηλαδή να γράψουμε $R \cap S = S \cap R$ και $R \cup S = S \cup R$, κάτι όμως που δεν ισχύει για τη διαίρεση, όπου είναι $R - S \neq S - R$.

Προκειμένου να κατανοήσουμε καλύτερα τον τρόπο λειτουργίας αυτών των τριών πράξεων, ας θεωρήσουμε τους πίνακες **STUDENT** και **INSTRUCTOR**.

STUDENT	
FIRSTNAME	LASTNAME
Susan	Yao
Ramesh	Shah
Johnny	Kohler
Barbara	Jones
Amy	Ford
Jimmy	Wang
Ernest	Gilbert

FNAME	LNAME
John	Smith
Ricardo	Browne
Susan	Yao
Francis	Johnson
Ramesh	Shah

Η εφαρμογή των πράξεων της **ένωσης**, της **τομής** και της **αφαίρεσης** πάνω σε αυτούς τους δύο πίνακες – οι οποίοι είναι **συμβατοί ως προς την ένωση** έτσι ώστε να μπορούν να εφαρμοσθούν οι εν λόγω πράξεις – παρουσιάζεται στη συνέχεια

FIRSTNAME	LASTNAME
Susan	Yao
Ramesh	Shah
Johnny	Kohler
Barbara	Jones
Amy	Ford
Jimmy	Wang
Ernest	Gilbert
John	Smith
Ricardo	Browne
Francis	Johnson

STUDENT $\cup$ INSTRUCTOR

FIRSTNAME	LASTNAME
Johnny	Kohler
Barbara	Jones
Amy	Ford
Jimmy	Wang
Ernest	Gilbert

STUDENT – INSTRUCTOR

FIRSTNAME	LASTNAME
John	Smith
Ricardo	Browne
Francis	Johnson

INSTRUCTOR – STUDENT

FIRSTNAME	LASTNAME
Susan	Yao
Ramesh	Shah

STUDENT $\cap$ INSTRUCTOR

Τέλος, το **καρτεσιανό γινόμενο (Cartesian product)** εφαρμόζεται ανάμεσα σε δύο πίνακες **R** και **S** – οι οποίοι στην προκειμένη περίπτωση **δεν είναι υποχρεωτικό να είναι συμβατοί ως προς την ένωση** – και δημιουργεί έναν τρίτο πίνακα που περιέχει **τους συνδυασμούς όλων των πλειάδων του ενός, με όλες τις πλειάδες του άλλου**. Με τον τρόπο αυτό είναι δυνατόν να εντοπίσουμε πλειάδες στους δύο πίνακες, οι οποίοι συσχετίζονται μεταξύ τους.

Στη γενική περίπτωση, το καρτεσιανό γινόμενο ανάμεσα σε δύο πίνακες **R(A₁, A₂, ..., A_n)** και **S(B₁, B₂, ..., B_m)** ορίζεται ως ένας καινούριος πίνακας **Q** που περιλαμβάνει συνολικά **m+n** πεδία και η δομή του είναι η {A₁, A₂, ..., A_n, B₁, B₂, ..., B_m}. Εάν ο πίνακας **R** περιέχει **M** πλειάδες και ο πίνακας **S** περιέχει **N** πλειάδες, ο πίνακας **Q = R X S** θα περιέχει συνολικά **M*N** πλειάδες.

Ας δούμε τώρα ένα παράδειγμα που να περιλαμβάνει τη συνδυασμένη χρήση των πράξεων της **επιλογής**, της **προβολής**, και του **καρτεσιανού γινομένου**. Έστω ότι θέλουμε για **κάθε γυναίκα υπάλληλο της εταιρείας να ανακτήσουμε τα ονόματα των προστατευόμενων μελών της**. Για να το κάνουμε αυτό, θα ακολουθήσουμε την επόμενη διαδικασία:

(α) Βρίσκουμε όλες τις γυναίκες υπαλλήλους της εταιρείας

$$\text{FEMALE_EMPS} \leftarrow \sigma_{\text{SEX}='F'}(\text{EMPLOYEE})$$

(β) Από όλα τα πεδία του πίνακα που θα προκύψει, κρατάμε μόνο το **LNAME**, το **FNAME**, και το **SSN**

$$\text{EMP_NAMES} \leftarrow \pi_{\text{LNAME, FNAME, SSN}}(\text{FEMALE_EMPS})$$

(γ) Επειδή τα ονόματα των προστατευόμενων μελών των υπαλλήλων της εταιρείας βρίσκονται στον πίνακα **DEPENDENT** θα πρέπει να συνδυάσουμε τους πίνακες **EMP_NAMES** και **DEPENDENT** έτσι ώστε να ανακτήσουμε όλους τους δυνατούς συνδυασμούς των πλειάδων των δύο πινάκων

$$\text{EMP_DEP} \leftarrow \text{EMP_NAMES} \bowtie \text{DEPENDENTS}$$

(δ) Για κάθε υπάλληλο, τα προστατευόμενα μέλη του βρίσκονται καταχωρημένα στις πλειάδες του πίνακα **DEPENDENT** για τις οποίες το πεδίο **ESSN** έχει την ίδια τιμή με το πεδίο **SSN** του θεωρούμενου υπαλλήλου. Επομένως από όλες τις πλειάδες που περιλαμβάνονται στο καρτεσιανό γινόμενο των δύο πινάκων, θα κρατήσουμε μόνο εκείνες για τις οποίες ισχύει η σχέση **ESSN = SSN**.

$$\text{ACTUAL_DEP} \leftarrow \sigma_{\text{SSN}=\text{ESSN}}(\text{EMP_DEP})$$

(ε) Τέλος από όλα τα πεδία των πλειάδων που ανακτήσαμε, κρατάμε μόνο το ονοματεπώνυμο του υπαλλήλου και τα ονόματα των προστατευόμενων μελών του:

$$\text{RESULT} \leftarrow \pi_{\text{LNAME, FNAME, DEP_NAME}}(\text{ACTUAL_DEP})$$

Από το παραπάνω παράδειγμα διαπιστώνουμε πως σε όλες σχεδόν τις περιπτώσεις η **πράξη του καρτεσιανού γινομένου ακολουθείται πάντα από την πράξη της επιλογής** προκειμένου να επιλέξουμε εκείνους τους συνδυασμούς πλειάδων που πληρούν κάποια συγκεκριμένη συνθήκη. Αυτό σημαίνει πως το να καλέσουμε το καρτεσιανό γινόμενο από μόνο του, δεν έχει κάποια πρακτική εφαρμογή, αφού το μόνο που κάνει είναι να επιστρέφει όλους τους δυνατούς συνδυασμούς των πλειάδων των δύο πινάκων. Εάν όμως από όλες αυτές τις πλειάδες κρατήσουμε μόνο όσες πληρούν κάποια συνθήκη, η ακολουθία αυτών των δύο πράξεων, έχει νόημα διότι ανακτούμε επιλεκτικά πληροφορία από δύο – ή περισσότερους – πίνακες της βάσης. Επειδή αυτή η ακολουθία **CARTESIAN PRODUCT + SELECT** χρησιμοποιείται πάρα πολύ συχνά, έχει ορισθεί μια νέα πράξη – η **πράξη της σύζευξης** – η οποία επι-

τρέπει την εκτέλεση όλης αυτής της διαδικασίας σε ένα και μόνο βήμα. Αυτή η πράξη αποτελεί το αντικείμενο της επόμενης ενότητας.

Η ΠΡΑΞΗ ΤΗΣ ΣΥΖΕΥΞΗΣ (JOIN OPERATION)

Όπως έχει αναφερθεί στην προηγούμενη παράγραφο, η πράξη της σύζευξης μπορεί να ορισθεί ως η εφαρμογή της πράξης του καρτεσιανού γινομένου ανάμεσα σε δύο πίνακες R και S , συνοδευόμενη από την πράξη της επιλογής (select), που επιστρέφει, όχι όλους τους συνδυασμούς πλειάδων των δύο πινάκων, αλλά μόνο εκείνους που ικανοποιούν κάποια συγκεκριμένη συνθήκη επιλογής. Αυτό σημαίνει πως για να ορίσουμε την πράξη της σύζευξης, χρειαζόμαστε δύο πίνακες και μια συνθήκη. Στην περίπτωση αυτή μπορούμε να γράψουμε

$$\text{RESULT} \leftarrow S \bowtie_{\langle \text{selection condition} \rangle} R$$

σύνταξη, η οποία είναι ισοδύναμη με την ακολουθία πράξεων

$$\sigma_{\langle \text{selection condition} \rangle} (S \times R)$$

Έτσι, στο προηγούμενο παράδειγμα, οι πράξεις

$$\text{EMP_DEP} \leftarrow \text{EMP_NAMES} \times \text{DEPENDENTS}$$

$$\text{ACTUAL_DEP} \leftarrow \sigma_{\text{SSN=ESSN}} (\text{EMP_DEP})$$

μπορούν να αντικατασταθούν από τη σύζευξη

$$\text{ACTUAL_DEP} \leftarrow \text{EMP_NAMES} \bowtie_{\text{SSN=ESSN}} \text{DEPENDENTS}$$

Το αποτέλεσμα της πράξης της σύζευξης ανάμεσα σε δύο πίνακες $R(A_1, A_2, \dots, A_n)$ και $S(B_1, B_2, \dots, B_m)$, είναι ένας καινούριος πίνακας Q με δομή $\{A_1, A_2, \dots, A_n, B_1, B_2, \dots, B_m\}$ που περιλαμβάνει μόνο τις πλειάδες του καρτεσιανού γινομένου οι οποίες πληρούν τη συνθήκη επιλογής που χρησιμοποιείται. Αυτή άλλωστε είναι και η βασική διαφορά που υφίσταται ανάμεσα στις πράξεις της σύζευξης και του καρτεσιανού γινομένου: στο καρτεσιανό γινόμενο περιλαμβάνονται όλοι οι δυνατοί συνδυασμοί των πλειάδων των δύο πινάκων, ενώ στο αποτέλεσμα της σύζευξης υπάρχουν μόνο εκείνες οι πλειάδες οι οποίες πληρούν τη συνθήκη επιλογής. Αυτή η συνθήκη επιλογής – όπως έχουμε ήδη αναφέρει στην εντολή **SELECT** – εφαρμόζεται σε κάθε πλειάδα του καρτεσιανού γινομένου ξεχωριστά, και στο τελικό αποτέλεσμα, προστίθενται μόνο εκείνες οι εγγραφές για τις οποίες ισχύει. Ας σημειωθεί πως αυτή η συνθήκη επιλογής μπορεί να είναι τόσο απλή όσο και σύνθετη, και να περιλαμβάνει τόσο τελεστές σύγκρισης όσο και λογικούς τελεστές.

Στην περίπτωση κατά την οποία η συνθήκη επιλογής πλειάδων από το καρτεσιανό γινόμενο των δύο πινάκων, περιλαμβάνει μόνο τον τελεστή της ισότητας (=) όπως συμβαίνει στο προηγούμενο παράδειγμα, η πράξη της σύζευξης ονομάζεται

EQUIJOIN. Στην περίπτωση αυτή, στο τελικό αποτέλεσμα, θα υπάρχουν δύο στήλες με διαφορετικά ονόματα που προέρχονται από τους δύο πίνακες – μία στήλη από κάθε πίνακα – αλλά έχουν ακριβώς τα ίδια αποτελέσματα. Για παράδειγμα, εάν θεωρήσουμε την σύζευξη

DEPT_MGR $\leftarrow$ **DEPARTMENT** $\bowtie$ MGRSSN=SSN **EMPLOYEE**

η οποία για κάθε **DEPARTMENT** επιστρέφει το **MANAGER** που το διευθύνει, ο πίνακας **DEPT_MGR** θα έχει την ακόλουθη δομή :

DEPT_MGR {DNAME, DNUMBER, MGRSSN, MGRSTARTDATE, FNAME, MINIT, LNAME, SSN, BDATE, ADDRESS, SEX, SALARY, SUPERSSN, DNO}

Εάν εκτυπώσουμε τα περιεχόμενα του πίνακα, θα διαπιστώσουμε πως οι στήλες **SSN** και **MGRSSN** έχουν ακριβώς τα ίδια περιεχόμενα, κάτι βέβαια είναι αναμενόμενο αφού ως συνθήκη επιλογής χρησιμοποιήσαμε την **MGRSSN = SSN**. Με άλλα λόγια, η πράξη της σύζευξης έτσι όπως την έχουμε ορίσει **επιστρέφει δυο στήλες οι οποίες είναι ακριβώς οι ίδιες**. Εάν θέλουμε στο αποτέλεσμα που θα λάβουμε να κρατήσουμε μόνο τη μια από τις δύο στήλες, καλούμε μια διαφορετική μορφή της σύζευξης που λέγεται **φυσική σύζευξη (natural join)** και κρατά μόνο τη μια από τις δύο ίδιες στήλες και πιο συγκεκριμένα εκείνη που βρίσκεται στο αριστερό μέρος της ισότητας που καθορίζουμε στη συνθήκη επιλογής. Η πράξη της φυσικής σύζευξης συμβολίζεται με το «*» και επειδή σε όλες τις περιπτώσεις προϋποθέτει τη χρήση του **τελεστή ισότητας (=)** στη συνθήκη επιλογής, αυτός μπορεί να παραλειφθεί. Έτσι, η προηγούμενη πράξη μπορεί τώρα να γραφεί με τη μορφή

DEPT_MGR $\leftarrow$ **DEPARTMENT** * $(\text{MGRSSN}), (\text{SSN})$ **EMPLOYEE**

με το τελικό αποτέλεσμα να περιλαμβάνει μόνο τη στήλη **MGRSSN**, ενώ, για να κρατήσουμε τη στήλη **SSN**, θα πρέπει να γράψουμε

DEPT_MGR $\leftarrow$ **EMPLOYEE** * $(\text{SSN}), (\text{MGRSSN})$ **DEPARTMENT**

Τέλος εάν τα δύο πεδία στη συνθήκη επιλογής έχουν το ίδιο όνομα, μπορούμε να τα παραλείψουμε τελείως κατά τη σύνταξη της εντολής. Εάν για παράδειγμα το πεδίο **MGRSSN** του πίνακα **DEPARTMENT** είχε και αυτό το όνομα **SSN** θα μπορούσαμε απλά να γράψουμε

DEPT_MGR $\leftarrow$ **EMPLOYEE** * **DEPARTMENT**

Είναι προφανές, πως εάν καλέσουμε την πράξη της σύζευξης χωρίς να καθορίσουμε κάποια συνθήκη επιλογής, τότε αυτή θα εκφυλισθεί στην πράξη του καρτεσιανού γινομένου, και θα μας επιστρέψει όλους τους δυνατούς συνδυασμούς πλειάδων των δύο πινάκων.

OUTER JOIN ΚΑΙ OUTER UNION

Εκτός από τις πράξεις της σύζευξης και της ένωσης που ορίσαμε στις προηγούμενες παραγράφους, υπάρχουν και κάποιες παραλλαγές τους οι οποίες δεν αποτελούν τμήμα της σχεσιακής άλγεβρας, αλλά έχουν υποστηρίζονται από όλα τα μοντέρνα σχεσιακά συστήματα διαχείρισης βάσεων δεδομένων. Αυτές οι πράξεις είναι η **εξωτερική σύζευξη (outer join)** και η **εξωτερική ένωση (outer union)**.

Η διαφορά της εξωτερικής σύζευξης από την απλή σύζευξη, έχει να κάνει με το πλήθος του συνδυασμού των πλειάδων των δύο πινάκων που επιστρέφονται. Στην απλή σύζευξη, όπως είναι γνωστό, επιστρέφονται **μόνο οι συνδυασμοί εκείνων των πλειάδων τα στοιχεία των οποίων ικανοποιούν τη συνθήκη επιλογής που καθορίζουμε στη σύζευξη**. Αντίθετα, στην πράξη της εξωτερικής σύζευξης, **διατηρούνται όλες οι πλειάδες, ακόμη και εκείνες που δεν πληρούν τον κανόνα επιλογής**. Αυτές οι πλειάδες, συνδυάζονται με «κενές» πλειάδες του άλλου πίνακα, δηλαδή με πλειάδες που έχουν σε όλα τα πεδία τους την τιμή **NULL**. Η πράξη της εξωτερικής σύζευξης μπορεί να είναι τόσο **αριστερή (left outer join)** όσο και **δεξιά (right outer join)** και ένα από τα σύμβολα που χρησιμοποιούνται για την αναπαράστασή της είναι το (+).

Ας υποθέσουμε για παράδειγμα πως θέλουμε να ανακτήσουμε **τα ονόματα των υπαλλήλων της εταιρείας**, καθώς και **τα ονόματα των τμημάτων που διευθύνουν** – εάν τυχαίνει να διευθύνουν κάποιο τμήμα. Στην περίπτωση αυτή, ο πίνακας που θα περιέχει το αποτέλεσμα του ερωτήματος, θα έχει τη δομή {FNAME, MINIT, LNAME, DNAME} με το πεδίο **DNAME** να περιέχει το όνομα του τμήματος, για όσους υπαλλήλους διευθύνουν κάποιο τμήμα, και την τιμή **NULL** για όλους τους υπόλοιπους. Προκειμένου να απαντήσουμε σε αυτό το ερώτημα, θα πρέπει να πραγματοποιήσουμε την επόμενη ακολουθία εντολών

TEMP ← EMPLOYEE (+) MGRSSN=SSN DEPARTMENT

RESULT ← $\pi_{\text{LNAME, MINIT, FNAME, DNAME}}$ (TEMP)

η οποία θα μας δώσει τα ακόλουθα αποτελέσματα

FNAME	MINIT	LNAME	DNAME
John	B	Smith	NULL
Franklin	T	Wong	Research
Alicia	J	Zelaya	NULL
Jennifer	S	Wallace	Administration
Ramesh	K	Narayan	NULL
Joyce	A	English	NULL
Ahmad	V	Jabbar	NULL
James	E	Borg	Headquarters

Ας περάσουμε τώρα στην πράξη της **εξωτερικής ένωσης (outer union)**. Η διαφορά που υφίσταται ανάμεσα στην εξωτερική ένωση και στην πράξη της ένωσης

πινάκων που μελετήσαμε σε προηγούμενες παραγράφους, είναι πως η εξωτερική ένωση ανάμεσα σε δύο πίνακες μπορεί να πραγματοποιηθεί, ακόμη και εάν οι δύο πίνακες δεν είναι συμβατοί ως προς την ένωση (**union compatible**). Με άλλα λόγια η προϋπόθεση της συμβατότητας ως προς την ένωση, δεν ισχύει για όλα τα πεδία των πινάκων αλλά μόνο για μερικά από αυτά – στην περίπτωση αυτή λέμε πως οι δύο πίνακες είναι μερικώς συμβατοί ως προς την ένωση (**partially union compatible**). Ένα παράδειγμα εφαρμογής αυτής της πράξης, είναι η εξωτερική ένωση των πινάκων **STUDENT** {Name, SSN, Department, Advisor} και **FACULTY** {Name, SSN, Department, Rank}. Στην περίπτωση αυτή το αποτέλεσμα αυτής της πράξης, θα είναι ένας πίνακας με δομή {Name, SSN, Department, Advisor, Rank} που θα περιέχει όλες τις πλειάδες των δύο πινάκων. Επειδή όμως στον πίνακα **STUDENT** δεν υπάρχει το πεδίο **rank**, η τιμή αυτού του πεδίου σε όλες τις εγγραφές που προέρχονται από αυτόν τον πίνακα, θα είναι η τιμή **NULL**. Με τον ίδιο τρόπο, σε όλες τις εγγραφές που προέρχονται από τον πίνακα **FACULTY**, το πεδίο **Advisor** θα έχει τιμή **NULL**, διότι αυτό το πεδίο δεν υπάρχει στον πίνακα **FACULTY**.

Η ΠΡΑΞΗ ΤΗΣ ΔΙΑΙΡΕΣΗΣ (DIVISION OPERATION)

Η πράξη της διαίρεσης εφαρμόζεται σε δύο πίνακες **A** και **B** εκ των οποίων ο **A** έχει περισσότερες στήλες από τον **B**, και επιστρέφει ένα νέο πίνακα που περιέχει τις στήλες του **A** που δεν ανήκουν στον **B**, και εκείνες τις τιμές που πίνακα **A** που συνδυάζονται με όλες τις τιμές του πίνακα **B**. Για την αναπαράσταση της πράξης της διαίρεσης, χρησιμοποιείται το σύμβολο «÷»

Για να κατανοήσουμε την πράξη της διαίρεσης ας θεωρήσουμε το ακόλουθο παράδειγμα. Έστω ότι θέλουμε να βρούμε τα ονόματα όλων των υπαλλήλων που δουλεύουν σε όλα τα **PROJECTS** στα οποία δουλεύει ο **John Smith**. Για να απαντήσουμε σε αυτό το ερώτημα, αρχικά θα πρέπει να ανακτήσουμε τη λίστα με τους κωδικούς των **PROJECTS** στα οποία δουλεύει ο **John Smith** – το αποτέλεσμα αυτού του ερωτήματος θα καταχωρηθεί στον πίνακα **SMITH_PROJ** {PNO}.

SMITH ← $\sigma_{\text{FNAME='JOHN' AND LNAME='SMITH'}}$ (**EMPLOYEE**)

SMITH_PROJ ← π_{PNO} (**WORKS_ON** *_{ESSN=SSN} **SMITH**)

Στο επόμενο βήμα της διαδικασίας, θα πρέπει να ανακτήσουμε για κάθε **EMPLOYEE** τον κωδικό του **SSN**, και τους κωδικούς των **PROJECTS** στα οποία απασχολείται, κάτι που μπορεί να γίνει πολύ απλά γράφοντας

SSN_PROJ ← $\pi_{\text{PNO, ESSN}}$ (**WORKS_ON**)

Από τους δύο πίνακες στους οποίους έχουμε καταλήξει, ο πίνακας **SSN_PROJ** {PNO, ESSN} περιλαμβάνει τους κωδικούς **SSN** όλων των **EMPLOYEES** που δουλεύουν σε κάποιο **PROJECT** καθώς και τον κωδικό του κάθε **PROJECT**, ενώ ο πίνακας **SMITH_PROJ** {PNO}, περιλαμβάνει τους κωδικούς των **PROJECTS** στα οποία δουλεύει ο **John Smith**. Για να βρούμε τώρα τους κωδικούς **ESSN** των **EMPLOYEES** που δουλεύουν σε όλα τα **PROJECTS** στα οποία

δουλεύει ο **John Smith**, θα εφαρμόσουμε την πράξη της διαίρεσης ανάμεσα στους δύο πίνακες :

$$\text{SSNS (SSN)} \longleftarrow \text{SSN_PROJ} \div \text{SMITH_PROJ}$$

Σύμφωνα με τον ορισμό της διαίρεσης πινάκων, η παραπάνω πράξη θα δημιουργήσει ένα πίνακα που θα έχει όλες τις στήλες του **SSN_PROJ** που δεν ανήκουν στον **SMITH_PROJ**. Επειδή η δομή των δύο πινάκων είναι **SSN_PROJ** {PNO, ESSN} και **SMITH_PROJ** {PNO}, είναι προφανές, πως ο πίνακας **SSNS** που θα επιστρέψει η πράξη της διαίρεσης θα έχει μόνο μια στήλη, την **ESSN** – στο παραπάνω παράδειγμα χρησιμοποιούμε τη σύνταξη **SSNS (SSN)** για να μετονομάσουμε την στήλη **ESSN** σε **SSN**. Ποιες τιμές όμως θα περιέχει αυτή η στήλη για τους κωδικούς **SSN**? Θα περιέχει μόνο εκείνους τους κωδικούς που συνδυάζονται με όλες τις τιμές **PNO** στον πίνακα **SMITH_PROJ**. Επειδή όμως ο πίνακας **SMITH_PROJ** {PNO} περιλαμβάνει μόνο τους κωδικούς των **PROJECTS** στους οποίους εργάζεται ο **John Smith**, η παραπάνω διαίρεση θα δώσει τους κωδικούς **SSN** των υπαλλήλων που εργάζονται σε όλα αυτά τα **PROJECTS**. Για να κατανοήσουμε καλύτερα αυτή τη διαδικασία, ας δούμε τα περιεχόμενα των πινάκων **SMITH_PROJ** και **SSN_PROJ** καθώς και της διαίρεσής τους, **SSNS**, σύμφωνα με τα δεδομένα που έχουμε καταχωρήσει στους πίνακες της βάσης δεδομένων της εταιρείας:

SSN_PROJ		SMITH_PROJ	
ESSN	PNO	PNO	
123456789	1	1	
123456789	2	2	
666884444	3		
453453453	1		
453453453	2		
333445555	2		
333445555	3		
333445555	10		
333445555	20		
999887777	30		
999887777	10		
987987987	10		
987987987	30		
987654321	30		
987654321	20		
888665555	20		

SSNS	
SSN	
123456789	
453453453	

Εάν παρατηρήσουμε προσεκτικά τα περιεχόμενα αυτών των πινάκων θα διαπιστώσουμε πως υπάρχουν μόνο δύο κωδικοί **ESSN** στον πίνακα **SSN_PROJ** που να συνδυάζονται τόσο με το **PNO = 1** όσο και με το **PNO = 2**. Αυτοί οι δύο κωδικοί είναι ο **123456789** και ο **453453453**. Επειδή όμως οι κωδικοί **PNO = 1** και **PNO = 2** είναι οι κωδικοί των **PROJECTS** στα οποία εργάζεται ο **Smith**, έπεται πως οι δύο παραπάνω κωδικοί **SSN** είναι αναφέρονται στους υπαλλήλους που εργάζονται σε όλα

αυτά τα **PROJECTS** – και προφανώς κάποιος από αυτούς τους δύο κωδικούς (ο 123456789) θα ανήκει στον **John Smith**, αφού και αυτός συμπεριλαμβάνεται ανάμεσα σε αυτούς τους υπαλλήλους.

Τέλος, για να πάρουμε τα ονόματα όλων αυτών των υπαλλήλων, δεν έχουμε παρά να γράψουμε

RESULT $\leftarrow \pi_{\text{FNAME, LNAME}} (\text{SNNS} * \text{EMPLOYEE})$

ΑΘΡΟΙΣΤΙΚΕΣ ΣΥΝΑΡΤΗΣΕΙΣ (AGGREGATE FUNCTIONS)

Οι **αθροιστικές συναρτήσεις** δεν αποτελούν μέρος της σχεσιακής άλγεβρας αλλά έχουν ορισθεί προκειμένου να πραγματοποιήσουμε κάποιες **μαθηματικές πράξεις** πάνω στα δεδομένα των πινάκων της βάσης. Οι πιο σημαντικές από αυτές τις πράξεις είναι η **καταμέτρηση των εγγραφών ενός πίνακα (COUNT)**, ο **υπολογισμός αθροίσματος (SUM)** και του **μέσου όρου των τιμών μιας στήλης (AVERAGE)**, καθώς και η **εύρεση της μέγιστης (MAXIMUM) και της ελάχιστης (MINIMUM) από αυτές τις τιμές**. Οι παραπάνω συναρτήσεις μπορούν να εφαρμοσθούν πάνω σε όλες τις γραμμές του πίνακα, αλλά εναλλακτικά έχουμε τη δυνατότητα να ομαδοποιήσουμε τις εγγραφές του πίνακα με βάση ένα ή περισσότερα πεδία και στη συνέχεια να εφαρμόσουμε αυτές τις συναρτήσεις, σε κάθε ομάδα ξεχωριστά.

Ο τρόπος με τον οποίο μπορούμε να καλέσουμε αυτές τις συναρτήσεις χαρακτηρίζεται από μια σύνταξη της μορφής

$\langle \text{grouping attributes} \rangle \mathcal{F}_{\langle \text{function list} \rangle} (\langle \text{relation name} \rangle)$

όπου $\langle \text{grouping attributes} \rangle$ είναι η λίστα των πεδίων του πίνακα ως προς τα οποία θα λάβει χώρα ομαδοποίηση των εγγραφών πριν την εφαρμογή των αθροιστικών συναρτήσεων, $\langle \text{function list} \rangle$ είναι η λίστα των αθροιστικών συναρτήσεων που θα χρησιμοποιηθούν, και $\langle \text{relation name} \rangle$ είναι το όνομα του πίνακα στα δεδομένα του οποίου θα εφαρμοσθούν οι εν λόγω συναρτήσεις. Κάθε στοιχείο της λίστας των αθροιστικών συναρτήσεων, είναι ένα ζεύγος της μορφής $(\langle \text{function name} \rangle, \langle \text{attribute name} \rangle)$ που καθορίζει ποια συνάρτηση θα χρησιμοποιηθεί και πάνω σε πιο πεδίο του πίνακα θα εφαρμοσθεί. Έτσι, η εντολή

R $\leftarrow \text{DNO } \mathcal{F}_{\text{COUNT SSN, AVERAGE SALARY}} (\text{EMPLOYEE})$

ομαδοποιεί τους υπαλλήλους της εταιρείας με κριτήριο τον κωδικό του τμήματος στο οποίο εργάζονται, και στη συνέχεια, για τους υπαλλήλους του κάθε τμήματος, καταμετρά τον αριθμό τους και υπολογίζει το μέσο όρο των μισθών τους.

Το αποτέλεσμα της παραπάνω εντολής, παρουσιάζεται στον ακόλουθο πίνακα.

DNO	EMP_NO	AVG_SAL
5	4	33250
4	3	31000
1	1	55000

Στην περίπτωση κατά την οποία δεν καθορίσουμε κάποιο πεδίο ως προς το οποίο θα λάβει χώρα η ομαδοποίηση των εγγραφών, το αποτέλεσμα της παραπάνω εντολής είναι ένας πίνακας με μια μόνο γραμμή, αφού οι συναρτήσεις που θα καθορίσουμε, θα εφαρμοσθούν σε όλες τις εγγραφές του πίνακα, και όχι σε κάποια ομάδα εγγραφών. Έτσι, η εντολή

$$R \leftarrow \mathcal{F}_{\text{COUNT SSN, AVERAGE SALARY}}(\text{EMPLOYEE})$$

θα δώσει ως αποτέλεσμα τον πίνακα που ακολουθεί.

EMP_NO	AVG_SAL
8	35125

ΠΑΡΑΔΕΙΓΜΑΤΑ ΕΦΑΡΜΟΓΗΣ ΤΩΝ ΤΕΛΕΣΤΩΝ ΤΗΣ ΣΧΕΣΙΑΚΗΣ ΑΛΓΕΒΡΑΣ

Στις σελίδες που ακολουθούν παρουσιάζονται υποδειγματικά λυμένες ασκήσεις που έχουν ως στόχο να συμβάλουν στην καλύτερη κατανόηση του τρόπου με τον οποίο εφαρμόζουμε τους τελεστές της σχεσιακής άλγεβρας πάνω στα περιεχόμενα των πινάκων μιας σχεσιακής βάσης δεδομένων.

ΑΣΚΗΣΗ 1: Να βρεθούν τα ονόματα και οι διευθύνσεις όλων των υπαλλήλων που εργάζονται στο **Research Department**

ΑΠΑΝΤΗΣΗ: (α) Από τον πίνακα **DEPARTMENT** κρατάμε μόνο τη στήλη που αναφέρεται στο **Research Department**

$$\text{RESEARCH} \leftarrow \sigma_{\text{DNAME}='RESEARCH'}(\text{DEPARTMENT})$$

(β) Εφαρμόζουμε την πράξη της **σύζευξης** ανάμεσα στους πίνακες **RESEARCH** και **EMPLOYEE** με συνθήκη επιλογής την **DNUMBER=DNO** προκειμένου να ανακτήσουμε όλους τους **EMPLOYEES** που δουλεύουν στο **Research Department**

$$\text{RES_EMP} \leftarrow \text{RESEARCH} \bowtie_{\text{DNUMBER=DNO}} \text{EMPLOYEE}$$

(γ) Τέλος, από τον πίνακα **RES_EMP**, κρατάμε μόνο τα ονόματα και τις διευθύνσεις αυτών των υπαλλήλων.

$$\text{RESULT} \leftarrow \pi_{\text{FNAME, LNAME, ADDRESS}} (\text{RES_EMP})$$

ΑΣΚΗΣΗ 2: Για κάθε **PROJECT** που πραγματοποιείται στο **Stafford** να βρεθεί ο κωδικός του (PNUMBER), ο κωδικός του τμήματος που το παρακολουθεί, καθώς και το επώνυμο, η διεύθυνση και η ημερομηνία γέννησης του **MANAGER** αυτού του τμήματος.

ΑΠΑΝΤΗΣΗ: (α) Αρχικά, βρίσκουμε όλα τα **PROJECTS** που πραγματοποιούνται στο **Stafford**.

$$\text{STAFFORD} \leftarrow \sigma_{\text{PLOCATION='STAFFORD'}} (\text{PROJECT})$$

(β) Στη συνέχεια βρίσκουμε τα τμήματα που ελέγχουν όλα τα **PROJECTS** που γίνονται στο **Stafford**

$$\text{DEPT_PROJ} \leftarrow \text{STAFFORD} \bowtie_{\text{DNUM=DNUMBER}} \text{DEPARTMENT}$$

(γ) Στο επόμενο βήμα βρίσκουμε τους **MANAGERS** των τμημάτων που ελέγχουν όλα τα **PROJECTS** που γίνονται στο **Stafford**

$$\text{MAN_DEP_PRO} \leftarrow \text{DEPT_PROJ} \bowtie_{\text{MGRSSN=SSN}} \text{EMPLOYEE}$$

(δ) Τέλος, από όλες τις πληροφορίες που περιέχει ο πίνακας **MAN_DEP_PROJ**, κρατάμε μόνο τον κωδικό του **PROJECT**, τον κωδικό του τμήματος που το ελέγχει, και το επώνυμο, τη διεύθυνση και την ημερομηνία γέννησης του **MANAGER** για αυτό το τμήμα

$$\text{RESULT} \leftarrow \pi_{\text{PNUMBER,DNUM,LNAME,ADDRESS,BDATE}} (\text{MAN_DEP_PRO})$$

ΑΣΚΗΣΗ 3: Να βρεθούν τα ονόματα των υπαλλήλων που εργάζονται σε όλα τα **projects** τα οποία παρακολουθούνται από το **DEPARTMENT** με κωδικό 5.

ΑΠΑΝΤΗΣΗ: (α) Αρχικά βρίσκουμε τους κωδικούς όλων των **PROJECTS** που πραγματοποιούνται από το **DEPARTMENT** με κωδικό αριθμό 5. Για να το κάνουμε αυτό χρησιμοποιούμε ένα συνδυασμό των πράξεων **SELECT** και **PROJECT** όπως φαίνεται στη συνέχεια:

$$\text{DEPT5_PROJS} \leftarrow \pi_{\text{PNUMBER}} (\sigma_{\text{DNUM=5}} (\text{EMPLOYEE}))$$

(β) Στη συνέχεια για κάθε υπάλληλο που εργάζεται σε κάποιο **PROJECT** ανακτούμε τον κωδικό του και τον κωδικό του τμήματος στο οποίο εργάζεται από τον πίνακα **WORKS_ON**

$$\text{EMP_PROJ} (\text{SSN, PNO}) \leftarrow \pi_{\text{ESSN, PNO}} (\text{WORKS_ON})$$

(γ) Οι κωδικοί των υπαλλήλων που εργάζονται σε όλα τα **PROJECTS** που παρακολουθούνται από το **DEPARTMENT** με κωδικό αριθμό **5**, είναι όσοι κωδικοί του πίνακα **EMP_PROJ** συνδυάζονται με όλους τους κωδικούς του πίνακα **DEPT5_PROJS**. Επομένως θα ανακτηθούν ως το αποτέλεσμα της διαίρεσης των δύο πινάκων, και θα καταχωρηθούν στον πίνακα **EMP_SSNS** ο οποίος προκύπτει ως

$$\text{EMP_SSNS} \leftarrow \text{EMP_PROJ} \div \text{DEPT5_PROJ}$$

(δ) Τέλος, για να ανακτήσουμε τα ονοματεπώνυμα όλων αυτών των υπαλλήλων θα εκτελέσουμε την πράξη της φυσικής σύζευξης ανάμεσα στους πίνακες **EMP_SSNS** και **EMPLOYEE** και από τον πίνακα που θα προκύψει, θα κρατήσουμε μόνο τα πεδία **LNAME** και **LNAME**:

$$\text{RESULT} \leftarrow \pi_{\text{LNAME, FNAME}} (\text{EMP_SSNS} * \text{EMPLOYEE})$$

ΑΣΚΗΣΗ 4: Να βρεθεί ο κωδικός των **PROJECT** στα οποία εργάζεται ο υπάλληλος με το επώνυμο **Smith**, τόσο ως **εργαζόμενος**, όσο και ως **MANAGER** κάποιου τμήματος.

ΑΠΑΝΤΗΣΗ: Για να απαντήσουμε σε αυτό το ερώτημα, θα πρέπει να βρούμε τους κωδικούς των **PROJECT** στα οποία ο **Smith** εργάζεται (1) ως **εργαζόμενος** και (2) ως **MANAGER** και στη συνέχεια να ενώσουμε τους δύο πίνακες σε ένα και μοναδικό πίνακα, που θα περιέχει τους κωδικούς των **PROJECTS** στα οποία εργάζεται ο **Smith**, τόσο ως **εργαζόμενος** όσο και ως **MANAGER**.

(α) Βρίσκουμε από τον πίνακα **EMPLOYEE** τον κωδικό **SSN** του **Smith**

$$\text{SMITHS(ESSN)} \leftarrow \pi_{\text{SSN}} (\sigma_{\text{LNAME}='SMITH'} (\text{EMPLOYEE}))$$

(β) Ανακτούμε τους κωδικούς των **PROJECTS** στα οποία εργάζεται ο **Smith** ως **EMPLOYEE**

$$\text{SMITH_WORKER} \leftarrow \pi_{\text{PNO}} (\text{WORKS_ON} * \text{SMITHS})$$

(γ) Βρίσκουμε όλους τους **EMPLOYEES** που δουλεύουν ως **MANAGER** σε οποιοδήποτε τμήμα, και για κάθε έναν από αυτούς, κρατάμε το επώνυμό του και τον κωδικό του τμήματος που διευθύνει.

$$\text{MGR} \leftarrow \pi_{\text{LNAME, DNUMBER}} (\text{EMPLOYEE} \bowtie_{\text{SSN}=\text{MRGSSN}} \text{DEPARTMENT})$$

(δ) Από όλες τις γραμμές του παραπάνω πίνακα κρατάμε μόνο εκείνη που αναφέρεται στον **Smith**

$$\text{SMITH_MGR} \leftarrow \sigma_{\text{LNAME}='SMITH'} (\text{MGR})$$

(ε) Από την παραπάνω γραμμή κρατάμε μόνο τον κωδικό του τμήματος που διευθύνει ο **John Smith**

$$\text{SMITH_DEPT} \leftarrow \pi_{\text{DNUMBER}}(\text{SMITH_MGR})$$

(στ) Στη συνέχεια βρίσκουμε τους κωδικούς των **PROJECTS** που πραγματοποιούνται από το τμήμα το οποίο διευθύνεται από τον **John Smith**

$$\text{SMITH_MGR} \leftarrow \pi_{\text{PNUMBER}}(\text{SMITH_DEPT} * \text{PROJECT})$$

(ζ) Έχοντας ανακτήσει τους κωδικούς των **PROJECTS** στα οποία ο **Smith** εργάζεται σαν **EMPLOYEE** καθώς και εκείνων στα οποία ο **Smith** εργάζεται ως **MANAGER**, δεν έχουμε παρά να ενώσουμε τους αντίστοιχους πίνακες προκειμένου να λάβουμε τους κωδικούς των **PROJECTS** στα οποία ο **Smith** εργάζεται είτε ως **EMPLOYEE** είτε ως **MANAGER**. Έτσι τελικά θα έχουμε

$$\text{RESULT} \leftarrow \text{SMITH_WORKER} \cup \text{SMITH_MGR}$$

ΑΣΚΗΣΗ 5: Να βρεθούν τα ονόματα των υπαλλήλων που δεν έχουν προστατευόμενα μέλη.

ΑΠΑΝΤΗΣΗ: Για να απαντήσουμε στο ερώτημα αυτό θα πρέπει να βρούμε το σύνολο των κωδικών των υπαλλήλων που έχουν προστατευόμενα μέλη, και στη συνέχεια να τους αφαιρέσουμε από το σύνολο των κωδικών όλων των υπαλλήλων.

(α) Αρχικά βρίσκουμε το σύνολο των κωδικών όλων των υπαλλήλων

$$\text{ALL_EMPS} \leftarrow \pi_{\text{SSN}}(\text{EMPLOYEE})$$

(β) Στη συνέχεια βρίσκουμε τους κωδικούς των υπαλλήλων που έχουν προστατευόμενα μέλη

$$\text{EMP_WITH_DEP} \leftarrow \pi_{\text{ESSN}}(\text{DEPENDENT})$$

(γ) Οι κωδικοί των υπαλλήλων που δεν έχουν προστατευόμενα μέλη, είναι αυτοί που απομένουν εάν αφαιρέσουμε τους κωδικούς των υπαλλήλων που έχουν προστατευόμενα μέλη από το σύνολο των κωδικών των υπαλλήλων

$$\text{EMP_WITHOUT_DEP} \leftarrow (\text{ALL_EMPS} - \text{EMP_WITH_DEP})$$

(δ) Τέλος ανακτούμε τα ονόματα και τα επώνυμα αυτών των υπαλλήλων

$$\text{RESULT} \leftarrow \pi_{\text{LNAME, FNAME}}(\text{EMP_WITHOUT_DEPS} * \text{EMPLOYEE})$$

ΑΣΚΗΣΗ 6: Να βρεθούν τα ονόματα των **MANAGERS** που έχουν τουλάχιστον ένα προστατευόμενο μέλος.

ΑΠΑΝΤΗΣΗ: (α) Αρχικά ανακτούμε τους κωδικούς **SSN** των **MANAGERS** όλων των τμημάτων

$$\text{MGRS (SSN)} \leftarrow \pi_{\text{MGRSSN}} (\text{DEPARTMENT})$$

(β) Στη συνέχεια διαβάζουμε τους κωδικούς **SSN** όλων των **EMPLOYEES** που έχουν προστατευόμενα μέλη

$$\text{EMP_WITH_DEP} \leftarrow \pi_{\text{ESSN}} (\text{DEPENDENT})$$

(γ) Επειδή και οι **MANAGERS** είναι **EMPLOYEES**, οι κωδικοί **SSN** όλων των **MANAGERS** που έχουν προστατευόμενα μέλη, θα ανήκουν και αυτοί στον πίνακα **EMP_WITH_DEP**. Επομένως, οι κωδικοί των **MANAGERS** που έχουν προστατευόμενα μέλη, θα ανήκουν, τόσο στον πίνακα **MGRS** όσο και στον πίνακα **EMP_WITH_DEP**, και επομένως για να τους ανακτήσουμε θα πάρουμε την τομή των δύο πινάκων:

$$\text{MGRS_WITH_DEP} \leftarrow \text{MGRS} \cap \text{EMP_WITH_DEP}$$

(δ) Τέλος ανακτούμε τα ονόματα και τα επώνυμα αυτών των **MANAGERS**

$$\text{RESULT} \leftarrow \pi_{\text{LNAME, FNAME}} (\text{MGRS_WITHOUT_DEPS} * \text{EMPLOYEE})$$

ΑΣΚΗΣΗ 7: Να βρεθούν τα ονόματα των υπαλλήλων που ανήκουν στο τμήμα με κωδικό **5**, και εργάζονται περισσότερες από **10** ώρες την εβδομάδα στον **PROJECT** με όνομα **ProductX**

ΑΠΑΝΤΗΣΗ: (α) Αρχικά ανακτούμε τους κωδικούς των υπαλλήλων που εργάζονται στο τμήμα με κωδικό **5**

$$\text{SSNS} \leftarrow \pi_{\text{SSN}} (\sigma_{\text{DNO}=5} (\text{EMPLOYEE}))$$

(β) Στη συνέχεια βρίσκουμε τον κωδικό του **PROJECT** με όνομα **ProductX**

$$\text{PRODX (PNO)} \leftarrow \pi_{\text{PNUMBER}} (\sigma_{\text{PNAME}='PRODUCTX'} (\text{PROJECT}))$$

(γ) Από τον πίνακα **WORKS_ON** βρίσκουμε κρατάμε μόνο τις εγγραφές που αφορούν το **PROJECT ProductX** και αναφέρονται στους υπαλλήλους όλων των τμημάτων της εταιρείας που απασχολούνται σε αυτό

$$\text{WORKS_PRODX} \leftarrow \text{PRODX} * \text{WORKS_ON}$$

(δ) Από τον πίνακα **WORKS_PRODX** κρατάμε μόνο εκείνες τις εγγραφές που αφορούν τους υπαλλήλους του τμήματος **5** που εργάζονται στο **PROJECT ProductX** περισσότερες από **10** ώρες την εβδομάδα.

$$\text{EMP_PRODX} \leftarrow \sigma_{\text{HOURS} > 10}(\text{WORKS_PRODX} * \text{SSNS})$$

(ε) Τέλος βρίσκουμε το όνομα και τον επώνυμο για όλους αυτούς τους υπαλλήλους.

$$\text{RESULT} \leftarrow \pi_{\text{LNAME}, \text{FNAME}}(\text{EMP_PRODX} * \text{EMPLOYEE})$$

ΑΣΚΗΣΗ 8: Να βρεθούν τα ονόματα των υπαλλήλων που έχουν κάποιο προστατευόμενο μέλος, το μικρό όνομα του οποίου είναι ίδιο με το δικό τους.

ΑΠΑΝΤΗΣΗ: (α) Αρχικά ανακτούμε τους κωδικούς των υπαλλήλων που έχουν προστατευόμενα μέλη

$$\text{EMP_WITH_DEP} \leftarrow \pi_{\text{ESSN}}(\text{DEPENDENT})$$

(β) Στη συνέχεια ανακτούμε τα μικρά ονόματα αυτών των υπαλλήλων

$$\text{EMP_NAMES} \leftarrow \pi_{\text{FNAME}}(\text{EMP_WITH_DEP} * \text{EMPLOYEE})$$

(γ) Κατόπιν βρίσκουμε τους υπαλλήλους που έχουν το ίδιο μικρό όνομα με κάποιο από τα προστατευόμενα μέλη του

$$\text{EMP_DEP} \leftarrow \text{EMP_NAMES} \bowtie_{\text{FNAME}=\text{DEPENDENT_NAME}} \text{DEPENDENT}$$

(δ) Ανακτούμε τους κωδικούς αυτών των υπαλλήλων

$$\text{EMP_DEP_SSN} \leftarrow \pi_{\text{ESSN}}(\text{EMP_DEP})$$

(ε) Ανακτούμε το όνομα και το επώνυμο αυτών των υπαλλήλων

$$\text{RESULT} \leftarrow \pi_{\text{LNAME}, \text{FNAME}}(\text{EMP_DEP_SSN} * \text{EMPLOYEE})$$

ΑΣΚΗΣΗ 9: Να βρεθούν τα ονόματα των υπαλλήλων που εποπτεύονται απευθείας από τον **Franklin Wong**

ΑΠΑΝΤΗΣΗ: (α) Βρίσκουμε τον κωδικό SSN του **Franklin Wong**

$$\text{WONG_SSN} \leftarrow \sigma_{\text{FNAME}='FRANKLIN' \text{ AND } \text{LNAME}='WONG'}(\text{EMPLOYEE})$$

(β) Βρίσκουμε τους υπαλλήλους που εποπτεύονται απευθείας από τον **Franklin Wong**

$$\text{WONG_EMP} \leftarrow \text{WONG_SSN} \bowtie_{\text{SSN=SUPERSSN}} \text{EMPLOYEE}$$

(γ) Κρατάμε μόνο τα ονόματα και τα επώνυμα αυτών των υπαλλήλων

$$\text{RESULT} \leftarrow \pi_{\text{LNAME, FNAME}}(\text{WONG_EMP})$$

ΑΣΚΗΣΗ 10: Για κάθε **PROJECT** να βρεθεί το όνομα του **PROJECT** και ο συνολικός αριθμός των ωρών ανά εβδομάδα που αναλώνεται σε αυτό το **PROJECT** από όλους τους υπαλλήλους

ΑΠΑΝΤΗΣΗ: (α) Στο πρώτο βήμα θα ομαδοποιήσουμε τις εγγραφές του πίνακα **WORKS_ON** ως προς τον κωδικό του **project** και για κάθε ένα από τα **projects** της εταιρείας, θα αθροίσουμε τις ώρες που αφιερώνονται σε αυτό, από όλους τους υπαλλήλους της εταιρείας.

$$\text{PROJ_SUM}(\text{PNUMBER}, \text{HOURS}) \leftarrow \text{PNO} \mathcal{F}_{\text{SUM HOURS}}(\text{WORKS_ON})$$

(β) Τέλος για κάθε **project** ανακτούμε από τον πίνακα **PROJECT** το όνομά του, και το επιστρέφουμε μαζί με το συνολικό πλήθος των ωρών ανά εβδομάδα που απασχολούνται σε αυτό όλοι οι υπάλληλοι της εταιρείας.

$$\text{RESULT} \leftarrow \pi_{\text{PNAME, HOURS}}(\text{PROJ_NUM} * \text{PROJECT})$$

ΑΣΚΗΣΗ 11: Να ανακτηθούν τα ονόματα των υπαλλήλων που δουλεύουν σε όλα τα projects

ΑΠΑΝΤΗΣΗ: (α) Αρχικά ανακτούμε τους κωδικούς όλων των projects που πραγματοποιούνται από την εταιρεία.

$$\text{PROJ} \leftarrow \pi_{\text{PNUMBER}}(\text{PROJECT})$$

(β) Στη συνέχεια, από τον πίνακα **WORKS_ON** ανακτούμε για κάθε υπάλληλο, τον κωδικό του, καθώς και τον κωδικό των projects στα οποία απασχολείται.

$$\text{SSN_PROJ} \leftarrow \pi_{\text{PNO, ESSN}}(\text{WORKS_ON})$$

Οι υπάλληλοι που εργάζονται σε όλα τα **projects** έχουν κωδικούς στον πίνακα **SSN_PROJ**, οι οποίοι συνδυάζονται με όλους των κωδικούς των **projects** που βρίσκονται στον πίνακα **PROJ**. Επομένως για να ανακτήσουμε τους κωδικούς αυτών των υπαλλήλων, θα πρέπει να διαιρέσουμε τους δύο παραπάνω πίνακες. Έτσι θα έχουμε

$$\text{SSNS (SSN)} \leftarrow \text{SSN_PROJ} \div \text{PROJ}$$

(δ) Τέλος, ανακτούμε τα ονόματα και τα επώνυμα αυτών των υπαλλήλων γράφοντας

$$\text{RESULT} \leftarrow \pi_{\text{FNAME, LNAME}} (\text{SNNS} * \text{EMPLOYEE})$$

ΑΣΚΗΣΗ 12: Να ανακτηθούν τα ονόματα των υπαλλήλων που δεν εργάζονται σε κανένα **project**

ΑΠΑΝΤΗΣΗ: Οι κωδικοί των υπαλλήλων που δεν εργάζονται σε κανένα **project** προκύπτουν αν ανακτήσουμε τους κωδικούς των υπαλλήλων που εργάζονται στα διάφορα **projects** της εταιρείας, και τους αφαιρέσουμε από το σύνολο των κωδικών όλων των υπαλλήλων. Πιο αναλυτικά, έχουμε

(α) Βρίσκουμε τους κωδικούς των υπαλλήλων που εργάζονται στα διάφορα **projects** της εταιρείας.

$$\text{WORKER_SSN (SSN)} \leftarrow \pi_{\text{ESSN}} (\text{WORKS_ON})$$

(β) Βρίσκουμε τους κωδικούς όλων των υπαλλήλων της εταιρείας

$$\text{SSN (SSN)} \leftarrow \pi_{\text{SSN}} (\text{EMPLOYEE})$$

(γ) Για να βρούμε τους κωδικούς των υπαλλήλων που δεν εργάζονται σε κανένα **project** αφαιρούμε από το σύνολο των κωδικών των υπαλλήλων τους κωδικούς εκείνων των υπαλλήλων που εργάζονται στα διάφορα **projects**.

$$\text{NO_WORKER_SSN} \leftarrow \text{SSN} - \text{WORKER_SSN}$$

(δ) Τέλος, βρίσκουμε το όνομα και το επώνυμο αυτών των υπαλλήλων, γράφοντας

$$\text{RESULT} \leftarrow \pi_{\text{FNAME, LNAME}} (\text{NO_WORKER_SSN} * \text{EMPLOYEE})$$

ΑΣΚΗΣΗ 13: Για κάθε τμήμα να ανακτηθεί το όνομά του και ο μέσος όρος των μισθών για τους υπαλλήλους που δουλεύουν σε αυτό το τμήμα

ΑΠΑΝΤΗΣΗ: (α) Αρχικά βρίσκουμε τους υπαλλήλους που δουλεύουν σε όλα τα τμήματα της εταιρείας

$$\text{EMP_DEP} \leftarrow \text{EMPLOYEE} \bowtie_{\text{DNO=DNUMBER}} \text{DEPARTMENT}$$

(β) Στη συνέχεια ομαδοποιούμε τους υπαλλήλους με κριτήριο το τμήμα στο οποίο εργάζονται και υπολογίζουμε για κάθε τμήμα, το μέσο όρο των μισθών των υπαλλήλων που δουλεύουν σε αυτό.

DEP_AVG (DNUMBER, AVG) $\leftarrow$ $\mathcal{F}_{\text{AVERAGE SALARY}}(\text{EMP_DEP})$

(γ) Τέλος διαβάζουμε το όνομα του τμήματος και το επιστρέψουμε μαζί με το μέσο όρο των μισθών των υπαλλήλων που υπολογίσαμε παραπάνω.

RESULT $\leftarrow$ $\mathcal{P}_{\text{DNAME, AVG}}(\text{DEP_AVG} * \text{DEPARTMENT})$

ΑΣΚΗΣΗ 14: Να ανακτηθεί ο μέσος όρος των μισθών για όλες τις γυναίκες υπαλλήλους

ΑΠΑΝΤΗΣΗ: (α) Αρχικά βρίσκουμε όλες τις γυναίκες υπαλλήλους

FEMALE_EMPS $\leftarrow$ $\sigma_{\text{SEX}='F'}(\text{EMPLOYEE})$

(β) Στη συνέχεια επιστρέφουμε τον μέσο όρο των μισθών τους

AVG_FEM_SALARY $\leftarrow$ $\mathcal{F}_{\text{AVERAGE SALARY}}(\text{FEMALE_EMPS})$