



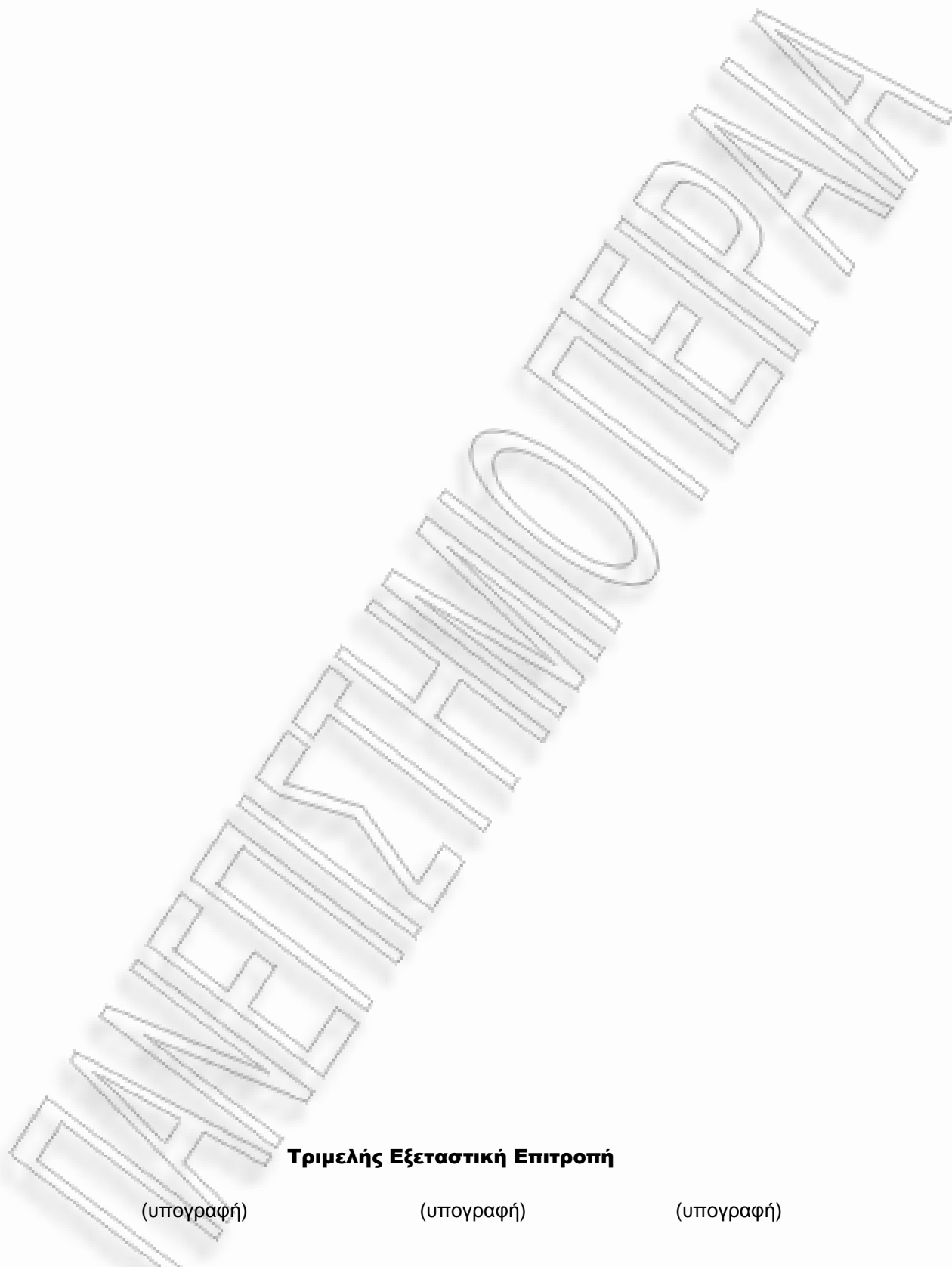
Πανεπιστήμιο Πειραιώς – Τμήμα Πληροφορικής
Πρόγραμμα Μεταπτυχιακών Σπουδών
«Πληροφορική»

Μεταπτυχιακή Διατριβή

Τίτλος Διατριβής	Συλλογή, Διασταύρωση, Διαχείριση και Επιχειρησιακή Αξιοποίηση Μεταναστευτικών Δεδομένων Ασφαλιστικών Οργανισμών
Όνοματεπώνυμο Φοιτητή	Σταυγιανουδάκης Πέτρος
Πατρώνυμο	Ιωάννης
Αριθμός Μητρώου	ΜΠΠΛ/ 06048
Επιβλέπων	Γιάννης Θεοδωρίδης, Αναπληρωτής Καθηγητής

Ημερομηνία Παράδοσης

Μάρτιος 2012



Τριμελής Εξεταστική Επιτροπή

(υπογραφή)

Γιάννης Θεοδωρίδης
Αναπλ. Καθηγητής

(υπογραφή)

Ιωάννης Σίσκος
Καθηγητής

(υπογραφή)

Νίκος Πελέκης
Λέκτορας

Περίληψη

Η εργασία σκιαγραφεί ένα σύστημα συλλογής, αποθήκευσης, επεξεργασίας και αξιοποίησης ασφαλιστικών δεδομένων που μπορεί να υποστηρίξει αποτελεσματικά τόσο τις διαδικασίες λήψης αποφάσεων για τη μεταναστευτική πολιτική, όσο και τον ευρύτερο ασφαλιστικό σχεδιασμό. Το σύστημα αξιοποιεί σύγχρονες τεχνολογίες (Βάσεις Δεδομένων, Data Warehouses, On-line Analytical Processing, GIS) για τη διαχείριση ασφαλιστικών στοιχείων και άλλων δημογραφικών, οικονομικών και λοιπών δεδομένων που αφορούν ασφαλισμένους μετανάστες ανά την ελληνική επικράτεια. Το παρόν κείμενο αποτελεί μια συνοπτική καταγραφή των τμημάτων της εργασίας που έχουν προς το παρόν υλοποιηθεί με την αξιοποίηση των ανωτέρω τεχνολογιών πληροφορικής.

Λέξεις κλειδιά: Βάσεις δεδομένων, Data Warehouses, OLAP, GIS.

Περίληψη	3
1.Εισαγωγή	6
2 Αποθήκες Δεδομένων	7
2.1 Γενικά	7
2.2 Αρχιτεκτονική συστημάτων αποθηκών δεδομένων	8
2.2.1 Πηγές και Μεταφορείς - Μετατροπείς	9
2.2.2 Αποθήκη Δεδομένων, Συλλογές Δεδομένων	10
2.2.3 Βάση Μετα-Δεδομένων.....	10
2.2.4 Σχεδίαση αρχιτεκτονικής Αποθηκών Δεδομένων.....	11
2.3 Μεταφορά δεδομένων από τις πηγές στην Αποθήκη Δεδομένων.....	11
2.3.1 Εξαγωγή και Μετατροπή Δεδομένων.....	12
2.3.2 Ολοκλήρωση	12
2.3.3 Εισαγωγή δεδομένων.....	12
2.3.4 Ενημέρωση	13
2.4 Σχεδίαση Αποθηκών Δεδομένων	13
2.5 Ειδικά θέματα των συστημάτων Αποθηκών Δεδομένων	16
2.5.1 Δομές Ευρετηρίων και η χρήση τους	16
2.5.2 Μετατροπή Πολύπλοκων Ερωτήσεων	16
2.6 Μοντέλα συστημάτων αναλυτικής επεξεργασίας	17
2.6.1 Πολυδιάστατα Μοντέλα Δεδομένων.....	17
2.6.2 Πράξεις στους υπερκύβους	19
3 Συστήματα υποστήριξης αποφάσεων και OLAP	21
3.1 Συστήματα Υποστήριξης Αποφάσεων	21
3.2 Ανάλυση Δεδομένων και OLAP	22
3.2.1 Online Αναλυτική Επεξεργασία	23
3.2.2 Υλοποίηση OLAP	27
3.2.3 Εκτεταμένες Συνοπτικές Λειτουργίες	28
3.2.4 Χαρακτηρισμοί.....	30
3.2.5 Παράθυρα	32
4. Γεωγραφικά Συστήματα Πληροφοριών	33
4.1 Ιστορικό	33
4.2 Θεμελιώδεις έννοιες & περιεχόμενο των Σ.Γ.Π. ως ΠΣ.	38
4.2.1 Ορισμοί - Θεμελιώδη Χαρακτηριστικά.....	38
4.2.2. Συστατικά Μέρη.....	39
4.3 Οργάνωση - Δόμηση	40
4.4 Χρήση - Τομείς Εφαρμογών	41
4.5 Χωρικά προβλήματα και χωρικός σχεδιασμός: Η οπτική των συστημάτων πληροφοριών	43
4.5.1 Τυπικά χωρικά προβλήματα και ΣΓΠ	43
4.5.2 ΣΓΠ και λήψη αποφάσεων	44
4.5.2.1 Ανάλυση συνθηκών – Αναγνώριση προβλημάτων	45
4.5.2.2 Σχεδιασμός λύσεων – Εναλλακτικά σενάρια	46
4.5.2.3 Επιλογή σεναρίων – λύσεων	46
4.6 Συνδυασμός τεχνολογιών G.I.S. και OLAP	47
5. Υλοποίηση συστήματος αναλυτικής επεξεργασίας μεταναστευτικών δεδομένων και οπτικοποίησης γεωγραφικής πληροφορίας	48
5.1 Αρχιτεκτονική συστήματος	48
5.2 Σχεδιασμός και υλοποίηση της Βάσης Δεδομένων του ΟΑΕΕ.....	49

5.2.1 Απαιτήσεις/ Περιορισμοί – ER.....	50
5.2.2 Σχεσιακό διάγραμμα – Ανάλυση βάσης δεδομένων	50
5.3 Αποθήκη Δεδομένων	52
5.3.1 Σχεδίαση της ΑΔ	54
5.4 Δημιουργία κύβου για OLAP ανάλυση.....	55
5.5 Δημιουργία χαρτών για οπτικοποίηση διαφόρων ερωτημάτων σε GIS	58
6. Αναλυτική επεξεργασία μεταναστευτικών δεδομένων και οπτικοποίηση γεωγραφικής πληροφορίας	62
6.1 Εύρεση πλήθους μεταναστών ανά Έτος και ανά Νομό διαμονής.....	62
6.2 Πλήθος ατόμων ανά Χώρα για το έτος 2008 για τους γομούς Δωδεκανήσου και Κυκλάδων	65
6.3 Πλήθος ατόμων ανά Έτος και ανά Νομό διαμονής και ανά Φύλο.....	67
6.4 Πλήθος ατόμων ανά Έτος και ανά Ήπειρο καταγωγής και ανά Οικογενειακή κατάσταση	68
6.5 Πλήθος ατόμων ανά Επαγγελματική δραστηριότητα και ανά Έτος	69
6.6 Ποσό εισφορών ανά χώρα καταγωγής	70
6.7 Πλήθος ατόμων ανά χώρα καταγωγής και ανά Οικονομική δραστηριότητα.....	72
6.8 Πλήθος ατόμων ανά φορέα και ανά χώρα γέννησης.....	73
6.9 Πλήθος ατόμων ανά δεκαετία γέννησης	73
6.10 Πλήθος ατόμων ανά χώρα καταγωγής και ανά Δεκαετία γέννησης.	74
6.11 Πλήθος ατόμων ανά Δεκαετία εγγραφής	75
6.12 Πλήθος ατόμων ανά Έτος και ανά Περιφέρεια Εργασίας.....	75
7 Βιβλιογραφικές αναφορές	77
8 ΠΑΡΑΡΤΗΜΑ	78
Δημιουργία Πινάκων βάσης δεδομένων.....	78
Δημιουργία Πινάκων Αποθήκης δεδομένων	82
Scripts εισαγωγής δεδομένων πίνακα μεταβολών σε Βάση Δεδομένων.....	84
ETL διαδικασία για εισαγωγή δεδομένων σε Αποθήκη Δεδομένων.....	88

1.Εισαγωγή

Αντικείμενο της εργασίας είναι η συλλογή δεδομένων από τους ασφαλιστικούς οργανισμούς (ηλικιακά, χώρες καταγωγής κλπ), η εισαγωγή αυτών σε βάσεις δεδομένων ανά ασφαλιστικό οργανισμό και ακολούθως η εκ νέου εισαγωγή των στοιχείων από τις βάσεις δεδομένων σε αποθήκη δεδομένων. Τελικό στάδιο είναι η υλοποίηση OLAP ερωτημάτων και η σύνδεση των αποτελεσμάτων των ερωτημάτων τους μέσω GIS με χάρτη.

Το όλο εγχείρημα ήταν μία εργασία του ΙΜΕΠΟ (Ίδρυμα Μεταναστευτικής Πολιτικής) για να λάβει αποφάσεις για το μέλλον της μεταναστευτικής πολιτικής.

Οι ασφαλιστικοί οργανισμοί από τους οποίους γίνεται η συλλογή των στοιχείων είναι το ΙΚΑ-ΕΤΑΜ (Ίδρυμα Κοινωνικών Ασφαλίσεων-Ενιαίο Ταμείο Ασφάλισης Μισθωτών) και ο ΟΑΕΕ/ΤΕΒΕ (Οργανισμός Ασφάλισης Ελεύθερων Επαγγελματιών / Ταμείο Επαγγελματιών και Βιοτεχνών Ελλάδος).

Το πρόγραμμα υλοποίησης των βάσεων δεδομένων και της αποθήκης δεδομένων είναι ο SQL SERVER 2005, του κύβου και των olap ερωτημάτων το Analysis Services 2005 και της οπτικοποίησης των ερωτημάτων σε χάρτη το Arcgis 9.1 .

Οι δύο βάσεις δεδομένων υλοποιήθηκαν ξεχωριστά από τους φοιτητές Σταυγιανουδάκη Πέτρο, για τον ΟΑΕΕ/ΤΕΒΕ, και την Καρασιώτου Βάσω, για το ΙΚΑ-ΕΤΑΜ. Ακολούθως με κατάλληλες ETL διαδικασίες έγινε φόρτωση των δεδομένων στην αποθήκη δεδομένων, όπου ορίστηκαν διαστάσεις και μέτρα στον κύβο

2 Αποθήκες Δεδομένων

2.1 Γενικά

Από τα μέσα της δεκαετίας του '70, η αλματώδης παραγωγή πολύ ισχυρών συστημάτων διαχείρισης βάσεων δεδομένων βοήθησε στην ανάπτυξη πληροφοριακών συστημάτων που καλύπτουν τις λειτουργικές ανάγκες οργανισμών και επιχειρήσεων. Τα μεγαλύτερα και ισχυρότερα συστήματα αναπτύχθηκαν με στόχο τον αυτοματισμό βασικών αναγκών των οργανισμών όπως η διεκπεραίωση των τραπεζικών εργασιών και τα λογιστικά συστήματα. Η λειτουργία αυτών των πληροφοριακών συστημάτων είναι πλέον κρίσιμη και πολύτιμη για τη ζωή των οργανισμών στους οποίους έχουν εγκατασταθεί, η δε βάση δεδομένων ενός τέτοιου συστήματος αποτελεί τον πυρήνα τους. Η ορθή σχεδίαση, ανάπτυξη και λειτουργία της βάσης είναι ο σημαντικότερος παράγοντας για την επιτυχία ενός πληροφοριακού συστήματος. Τα συστήματα αυτά παρέχουν τη δυνατότητα επεξεργασίας μεγάλου αριθμού δοσοληψιών που διαχειρίζονται τα δεδομένα του οργανισμού (On-line transaction processing - OLTP). Ένα άλλο είδος πληροφοριακών συστημάτων που αναπτύσσονται στους οργανισμούς είναι τα συστήματα στήριξης αποφάσεων που σκοπό έχουν να βοηθήσουν τα στελέχη των οργανισμών να σχεδιάσουν τις δραστηριότητές του. Η επιτυχία των συστημάτων αυτών είναι επίσης βασικός παράγοντας επιτυχίας του οργανισμού. Μία βασική απαίτηση των συστημάτων στήριξης αποφάσεων είναι η αποδοτική πρόσβαση στα δεδομένα των συστημάτων αυτοματισμού. Το πρόβλημα που προκύπτει, όμως, είναι ότι τα συστήματα αυτοματισμού έχουν ήδη πολύ σοβαρό υπολογιστικό φορτίο από μόνα τους και επιπλέον, είναι σχεδιασμένα για την εκτέλεση διαφορετικών λειτουργιών.

Ένας τηλεπικοινωνιακός οργανισμός, για παράδειγμα, συνήθως διαθέτει ένα μεγάλο πληροφοριακό σύστημα ελέγχου του τηλεφωνικού δικτύου του. Αυτό το σύστημα ελέγχει την ομαλή λειτουργία του δικτύου και παράλληλα των παροχή υπηρεσιών και την χρέωση των συνδρομητών του. Η βάση δεδομένων του συστήματος περιέχει όλα τα δεδομένα των παραπάνω εργασιών. Είναι σαφές ότι αυτό το σύστημα λειτουργεί συνεχώς (24 ώρες ημερησίως) με μεγάλο όγκο δοσοληψιών (transactions) να εξυπηρετούνται στη βάση δεδομένων. Από αυτή τη βάση θα πρέπει να αντλήσει και ένα σύστημα στήριξης αποφάσεων τα απαραίτητα δεδομένα, για να μπορέσει να βοηθήσει στο σχεδιασμό της λειτουργίας του οργανισμού. Μελετώντας λίγο πιο προσεκτικά την περίπτωση αυτή, θα δούμε ότι είναι πρακτικά αδύνατο το σύστημα ελέγχου του δικτύου και το σύστημα στήριξης αποφάσεων να λειτουργούν, χρησιμοποιώντας την ίδια βάση δεδομένων. Διάφορα προβλήματα κάνουν αδύνατη την εφαρμογή αυτού του σεναρίου. Τα κυριότερα από αυτά τα προβλήματα είναι τα παρακάτω:

Τα δύο συστήματα αναπτύχθηκαν πιθανότατα από διαφορετικούς ανθρώπους και κυρίως με τη χρήση διαφορετικών τεχνολογιών. Είναι πιθανό η τεχνολογία του συστήματος αποφάσεων να αδυνατεί να επιτρέψει άμεση πρόσβαση (on-line) στη βάση δεδομένων του συστήματος ελέγχου του δικτύου. Πολύ συχνά, σε μεγάλα συστήματα, όπως στην προκειμένη περίπτωση, το σύστημα ελέγχου του δικτύου έχει αναπτυχθεί με τη χρήση παρωχημένης τεχνολογίας, όπως, για παράδειγμα, αρχεία COBOL. Εφαρμογές που χρησιμοποιούν μοντέρνα τεχνολογία αντιμετωπίζουν προβλήματα στο να διαχειριστούν πληροφορία που προέρχεται από μια βάση δεδομένων παλαιάς τεχνολογίας.

Η βάση δεδομένων του συστήματος ελέγχου του δικτύου σχεδιάστηκε με βάση αποκλειστικά τις απαιτήσεις αυτής της εφαρμογής. Βασικό χαρακτηριστικό σε εφαρμογές αυτού του είδους είναι η όσο το δυνατό αποδοτικότερη ικανοποίηση μικρών δοσοληψιών που εισάγουν ή τροποποιούν πολύ μικρό αριθμό εγγραφών της βάσης. Μία τυπική δοσοληψία που θα αφορούσε τη χρέωση μίας υπεραστικής συνδιάλεξης θα εισήγαγε μία εγγραφή με τον κωδικό του συνδρομητή και τη διάρκεια της συνδιάλεξης. Στη σχεδίαση μιας τέτοιας βάσης δεδομένων, με την εφαρμογή των κανόνων κανονικοποίησης, καταλήγουμε σε μεγάλο αριθμό από πίνακες που ο κάθε ένας έχει περιορισμένο αριθμό πεδίων. Σε αντίθεση με τα παραπάνω, μία εφαρμογή που αντλεί στοιχεία λειτουργίας του δικτύου για λόγους ανάλυσης και λήψης αποφάσεων, δεν κάνει καμία αλλαγή στη βάση του δικτύου αλλά απαιτεί αποδοτική απόκριση από το σύστημα στις ερωτήσεις που θέτει. Αυτές οι ερωτήσεις συνήθως απαιτούν πρόσβαση σε μεγάλο αριθμό δεδομένων, θέτοντας διαφορετικούς κανόνες σχεδίασης της βάσης δεδομένων του συστήματος. Για μία ερώτηση σχετική με τη στρατηγική του

οργανισμού, που θα είχε πρόσβαση σε μεγάλο αριθμό δεδομένων το κόστος σε μία βάση με πολλούς πίνακες θα ήταν σημαντικό, καθώς θα έπρεπε να εκτελεστεί μεγάλο αριθμός από πράξεις JOIN μεταξύ των πινάκων αυτών.

Κάθε σύστημα στήριξης αποφάσεων που εκτελεί μεγάλο αριθμό ερωτήσεων, θα δεσμεύσει μεγάλο αριθμό πόρων του συστήματος διαχείρισης της βάσης δεδομένων με αποτέλεσμα να μειώσει την απόδοση του συστήματος ελέγχου του δικτύου. Για παράδειγμα, μία ερώτηση σχετική με τις χρεώσεις των πελατών του δικτύου, που απαιτεί κάποιο χρονικό διάστημα για να εκτελεστεί, θα κλείδωνε τον πίνακα με τις χρεώσεις των πελατών εμποδίζοντας οποιαδήποτε μεταβολή από το σύστημα ελέγχου (π.χ. μια νέα χρέωση).

Από τα παραπάνω γίνεται σαφές ότι είναι εξαιρετικά δυσχερής η χρήση των βάσεων δεδομένων των πληροφοριακών συστημάτων των οργανισμών από τα συστήματα στήριξης αποφάσεων. Όμως, η αποδοτική χρήση των συστημάτων στήριξης αποφάσεων απαιτεί όπως προαναφέρθηκε, πρόσβαση σε αυτά τα δεδομένα. Η εισαγωγή των “Αποθηκών Δεδομένων” είναι η λύση στο κρίσιμο αυτό πρόβλημα.

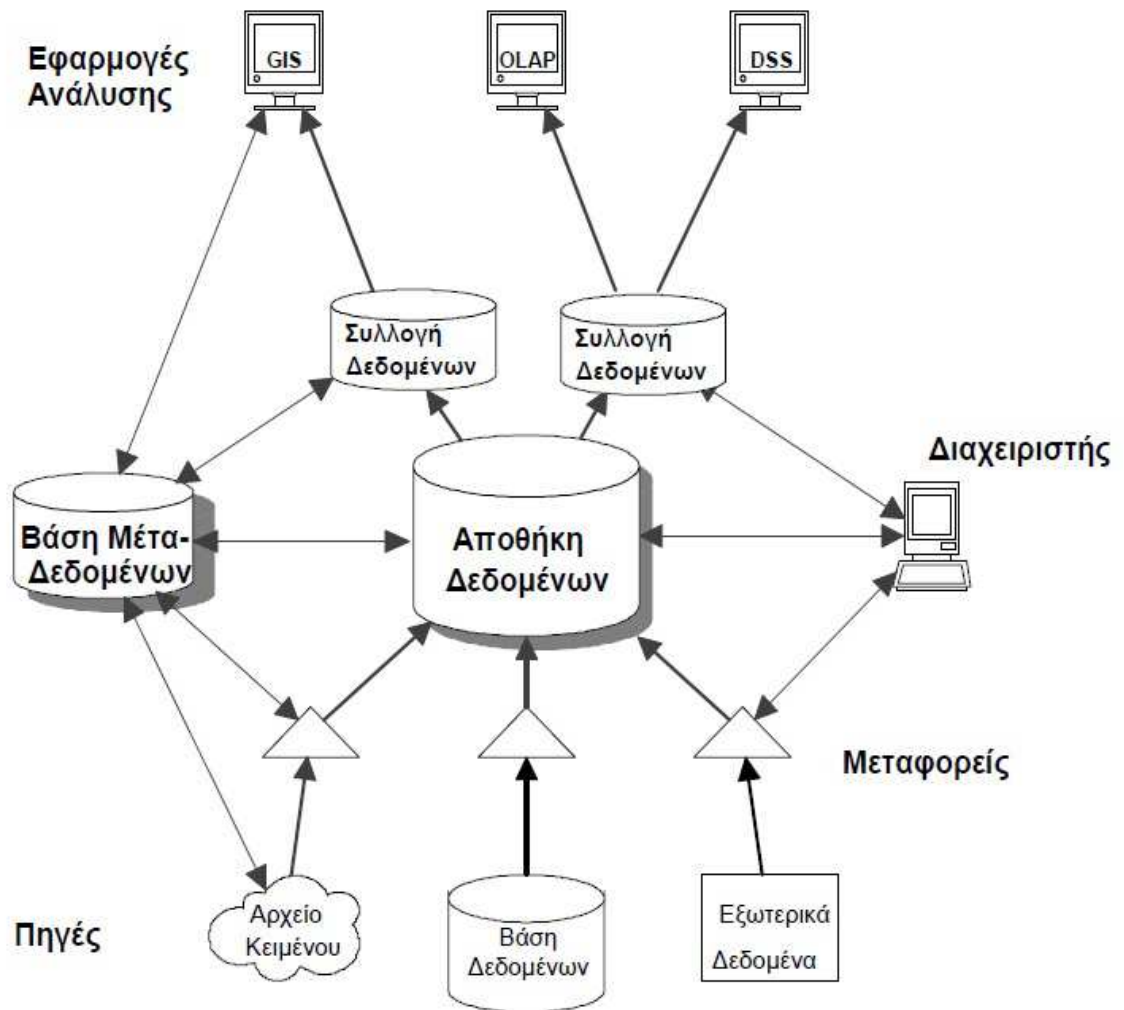
Με τον όρο *Αποθήκες Δεδομένων (Data Warehouses)* χαρακτηρίζουμε ένα σύνολο τεχνολογιών που επιτρέπει στους αναλυτές ενός οργανισμού στη σχεδίαση της πολιτικής του έχοντας αποδοτική πρόσβαση στα δεδομένα του οργανισμού. Μία Αποθήκη Δεδομένων διατηρεί δεδομένα που αντλεί από τις βάσεις δεδομένων των πληροφοριακών συστημάτων του οργανισμού αλλά και άλλες πηγές δεδομένων, όπως αρχεία του οργανισμού ή δεδομένα που προέρχονται από εξωτερικές πηγές. Αυτά τα δεδομένα οργανώνονται στην Αποθήκη Δεδομένων σε δομές κατάλληλες να απαντήσουν τις απαιτήσεις των αναλυτών - χρηστών των συστημάτων στήριξης αποφάσεων. Τα συστήματα στήριξης αποφάσεων αποκτούν πρόσβαση στα δεδομένα λειτουργίας του οργανισμού χωρίς την παρουσία των προαναφερθέντων προβλημάτων. Οι Αποθήκες Δεδομένων παρέχουν τη δυνατότητα για *Συνεχή Αναλυτική Επεξεργασία (On-Line Analytical Processing- OLAP)* των δεδομένων περιέχοντας συνήθως ιστορικά και συγκεντρωτικά δεδομένα που συνήθως αποδεικνύονται χρήσιμα για υποστήριξη αποφάσεων. Επίσης, παρέχουν μία ολοκληρωμένη εικόνα του σχήματος των δεδομένων του οργανισμού. Η σχεδίαση των Αποθηκών Δεδομένων έχει σαν στόχο την αποδοτική απάντηση των πολύπλοκων ερωτήσεων που θέτονται κατά την αναλυτική επεξεργασία δεδομένων από τις εφαρμογές στρατηγικού σχεδιασμού.

Η δημιουργία και η συντήρηση μίας Αποθήκης Δεδομένων είναι μία πολύπλοκη διαδικασία καθώς πολλές διαφορετικές προσεγγίσεις είναι εφικτές. Αρκετοί οργανισμοί επιδιώκουν να δημιουργήσουν μία Αποθήκη Δεδομένων που θα περιέχει αναλυτικά δεδομένα από όλες τις δραστηριότητες του οργανισμού. Πρόκειται για ένα πολύπλοκο εγχείρημα που απαιτεί μεγάλο κόστος για να επιτύχει. Μία άλλη λύση είναι η δημιουργία *Επιμέρους Συλλογών Δεδομένων (data marts)* με κριτήριο το αντικείμενο των εφαρμογών από τις οποίες προέρχονται ή το τμήμα του οργανισμού που τις χρησιμοποιεί. Πρόκειται για πιο ευέλικτα συστήματα στη δημιουργία τους, τα οποία όμως δεν παρέχουν ενιαία λύση, δημιουργώντας προβλήματα σε περίπτωση μακρόχρονης χρήσης τους.

Τα τελευταία χρόνια η ανάπτυξη και λειτουργία Αποθηκών Δεδομένων κρίνεται κρίσιμη για την λειτουργία των οργανισμών. Τεράστια ποσά επενδύονται σε αυτή τη δραστηριότητα ενώ τα οφέλη από τη λειτουργία τέτοιων συστημάτων κρίνονται ήδη ως ιδιαίτερα σημαντικά. Όπως είναι φυσικό, όλες οι μεγάλες εταιρείες του χώρου των Βάσεων Δεδομένων και των πληροφοριακών συστημάτων αναπτύσσουν και προτείνουν προϊόντα στο χώρο των Αποθηκών Δεδομένων. Τα επόμενα χρόνια αναμένονται ακόμα μεγαλύτερες επενδύσεις σε τεχνολογία αιχμής του χώρου.

2.2 Αρχιτεκτονική συστημάτων αποθηκών δεδομένων

Η επιλογή της αρχιτεκτονικής μιας αποθήκης δεδομένων πρέπει να ικανοποιεί τις συγκεκριμένες ανάγκες του οργανισμού για τις οποίες δημιουργήθηκε και να εξασφαλίζει τη διαθεσιμότητα και την αποδοτικότητα του συστήματος. Το Σχήμα 2.1 παρουσιάζει μια γενική αρχιτεκτονική ενός συστήματος Αποθήκης Δεδομένων. Στο σχήμα σημειώνονται τα βασικά δομικά στοιχεία μίας Αποθήκης Δεδομένων, η διασύνδεση των στοιχείων τους, καθώς και η ροή των δεδομένων.



Σχήμα 2.1 : Γενική Αρχιτεκτονική Αποθήκης Δεδομένων

Τα δομικά μέρη της αρχιτεκτονικής ενός συστήματος Αποθήκης Δεδομένων είναι τα ακόλουθα:

- Πηγές: Κάθε πηγή από την οποία η Αποθήκη Δεδομένων αντλεί δεδομένα.
- Μεταφορείς - Μετατροπείς: Εφαρμογές που εκτελούν τις διαδικασίες μεταφοράς των δεδομένων από τις πηγές στην Αποθήκη Δεδομένων.
- Αποθήκη Δεδομένων, Συλλογές Δεδομένων: Τα συστήματα που αποθηκεύονται τα δεδομένα που παρέχονται προς τους χρήστες.
- Βάση Μέτα-Δεδομένων: Σύστημα αποθήκευσης πληροφορίας σχετικά με τη δομή και λειτουργία του συστήματος.
- Διαχειριστής: Εφαρμογή που παρέχει δυνατότητα διαχείρισης του συστήματος
- Εφαρμογές Ανάλυσης: Εφαρμογές που έχουν πρόσβαση στην Αποθήκη Δεδομένων. Συνήθως είναι συστήματα στήριξης αποφάσεων.

2.2.1 Πηγές και Μεταφορείς - Μετατροπείς

Τα συστήματα διαχείρισης Αποθηκών Δεδομένων υποστηρίζουν άντληση δεδομένων

από διάφορες κατηγορίες πηγών δεδομένων. Οι συνηθέστερες από αυτές είναι:

- Βάσεις Δεδομένων των συστημάτων του οργανισμού.
- Εξωτερικές πηγές πληροφοριών, όπως για παράδειγμα, πληροφορίες που παρέχονται από πληροφοριακά συστήματα στα οποία υπάρχει πρόσβαση από τον οργανισμό.
- Αρχεία Εφαρμογών και αρχεία κειμένου.

Οι *Μεταφορείς / Μετατροπείς δεδομένων (wrappers / loaders)* είναι εφαρμογές που εξαγάγουν δεδομένα από τις πηγές και τα μεταφέρουν στην Αποθήκη Δεδομένων. Η ύπαρξη διαφορετικών κατηγοριών (Σχισιακές Βάσεις Δεδομένων, αρχεία COBOL, κείμενα MS-Word) που παρέχουν διαφορετική πρόσβαση στα δεδομένα τους οδηγεί στην ανάπτυξη διαφορετικών τύπων μεταφορών. Συνήθως, για κάθε μία διαφορετική πηγή, ή κατηγορία πηγής, ένας διαφορετικός μεταφορέας αναλαμβάνει να αντλεί τα δεδομένα της. Η λειτουργία αυτών των εφαρμογών κρίνεται ιδιαίτερα κρίσιμη για την επιτυχία του συστήματος, καθώς είναι υπεύθυνες για την αυτόματη μεταφορά, την επεξεργασία και τις αναγκαίες μετατροπές των δεδομένων από τις πηγές. Αναλυτικά, οι μεταφορείς αυτοματοποιούν τις παρακάτω διαδικασίες:

- Εξαγωγή δεδομένων από τις πηγές.
- Καθαρισμό των δεδομένων με την διάγνωση πιθανών ασυνεπειών και τη μεταφορά μόνο των πραγματικά χρήσιμων δεδομένων.
- Μετάδοση δεδομένων σε υψηλές ταχύτητες.
- Μετατροπή των δεδομένων μεταξύ διαφορετικών μοντέλων και προτύπων.
- Διάγνωση αλλαγών στα δεδομένα των πηγών και μεταφορά των νέων δεδομένων
- Εισαγωγή των δεδομένων στην Αποθήκη Δεδομένων.
- Δημιουργία αντιγράφων τμημάτων των πηγών στην Αποθήκη Δεδομένων.
- Ανάλυση των μεταφερόμενων δεδομένων για τη διάγνωση μη ορθής πληροφορίας.
- Έλεγχος πληρότητας Δεδομένων.

2.2.2 Αποθήκη Δεδομένων, Συλλογές Δεδομένων

Οι Αποθήκες Δεδομένων και οι Συλλογές Δεδομένων, όπως φαίνεται στο Σχήμα 2.1, υλοποιούνται με τη χρήση Σχισιακών Συστημάτων Διαχείρισης Βάσεων Δεδομένων. Τα δεδομένα αποθηκεύονται σε σχεσιακές βάσεις δεδομένων, ενώ πρόσβαση σε αυτά παρέχεται από μία γλώσσα διαχείρισης δεδομένων που είναι επέκταση της SQL. Εναλλακτική της χρήσης σχεσιακών συστημάτων είναι η χρήση των *Πολυδιάστατων Συστημάτων Αναλυτικής Επεξεργασίας (Multidimensional OLAP servers)*, που αποθηκεύουν και διαχειρίζονται δεδομένα με πολυδιάστατο τρόπο (βλ. και ενότητα 2.6). Η χρήση σχεσιακών ΣΔΒΔ εκμεταλλεύεται την ευελιξία και την ισχύ της τεχνολογίας των σύγχρονων συστημάτων. Κατά την αναλυτική επεξεργασία δεδομένων εκτελούνται πολύπλοκες ερωτήσεις που απαιτούν δυνατότητα διαχείρισης μεγάλου όγκου πληροφοριών. Τα πλεονεκτήματα των πολυδιάστατων συστημάτων βρίσκονται στη δυνατότητά τους να διαχειρίζονται δεδομένα, τα οποία είναι δομημένα με τρόπο που βρίσκεται πιο κοντά στις ανάγκες των εφαρμογών ανάλυσης (OLAP).

Η ύπαρξη των Συλλογών Δεδομένων είναι επιλογή του διαχειριστή του συστήματος. Οι Συλλογές Δεδομένων περιέχουν τμήματα των δεδομένων της Αποθήκης Δεδομένων. Ο καταμερισμός του περιεχομένου των Αποθηκών σε επιμέρους Συλλογές γίνεται με οργανωτικά κριτήρια και στόχο την πιο άμεση και αποδοτική πρόσβαση των εφαρμογών ανάλυσης στα δεδομένα της Αποθήκης καθώς και στον καταμερισμό των δεδομένων κατά αντικείμενο, ή τμήμα. Παράλληλα επιτυγχάνεται και η αποσυμφόρηση της Αποθήκης Δεδομένων.

2.2.3 Βάση Μετα-Δεδομένων

Τα *Μετα-Δεδομένα (metadata)*, έχουν ένα πολύ σημαντικό ρόλο στις Αποθήκες Δεδομένων. Η κατανόηση και η καταγραφή του περιεχομένου των δεδομένων και της οργάνωσής τους είναι απαραίτητη για τη αποδοτική λειτουργία και διαχείριση της Αποθήκης. Τα μετα-δεδομένα περιέχουν (ή καλύτερα, οφείλουν να περιέχουν):

- *Λεξικό Δεδομένων (Data Dictionary)* που περιέχει τον ορισμό και την περιγραφή των

δεδομένων που αποθηκεύονται στην Αποθήκη Δεδομένων και τις μεταξύ τους συσχετίσεις.

- Περιγραφή της ροής των δεδομένων μέσα στο σύστημα.
- Περιγραφή των κανόνων μετατροπής των δεδομένων κατά τη μεταφορά τους.
- Δεδομένα ελέγχου των διαφόρων εκδοχών (versions) των δεδομένων.
- Στατιστικά χρήσης των δεδομένων.
- Πληροφορία σχετικά με τους κανόνες ελέγχου πρόσβασης στην Αποθήκη Δεδομένων.
- Διάφορα ψευδώνυμα (aliases).

Όπως φαίνεται και στη Σχήμα 2.1, τα μετα-δεδομένα αποθηκεύονται σε ένα σύστημα, όπου υπάρχει πρόσβαση από κάθε δομικό στοιχείο της αρχιτεκτονικής. Το γεγονός αυτό δημιουργεί την ανάγκη ύπαρξης ενός σταθερού προτύπου για τα μετα-δεδομένα, καθώς, όπως προαναφέρθηκε, τα διάφορα δομικά στοιχεία που συμμετέχουν στην αρχιτεκτονική των Αποθηκών Δεδομένων είναι εφαρμογές ανεπτυγμένες ανεξάρτητα από τις πηγές και τις εφαρμογές ανάλυσης. Ένα τέτοιο πρότυπο έχει προταθεί από μια ομάδα εταιρειών του χώρου και ονομάζεται *Metadata Interchange Specification (MDIS)*.

Οι Αποθήκες Δεδομένων σε μερικές περιπτώσεις είναι καταμεμημένες ώστε να επιτυγχάνεται καταμερισμός του φορτίου, επεκτασιμότητα και διαθεσιμότητα του συστήματος. Σε αυτές τις καταμεμημένες αρχιτεκτονικές, υπάρχει συχνά ένα αντίγραφο του συστήματος των μετα-δεδομένων σε κάθε ένα από τους καταμεμημένους κόμβους της Αποθήκης, ενώ η όλη διαχείριση του συστήματος γίνεται από μία κεντρική εφαρμογή.

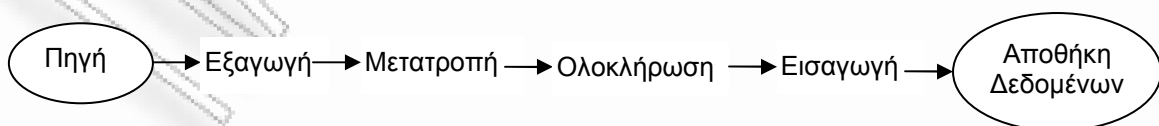
2.2.4 Σχεδίαση αρχιτεκτονικής Αποθηκών Δεδομένων

Η σχεδίαση μίας Αποθήκης Δεδομένων είναι μία πολύπλοκη διαδικασία που αποτελείται συνήθως από τις παρακάτω ενέργειες:

- Ορισμός της αρχιτεκτονικής και των απαιτούμενων στοιχείων του συστήματος. Επιλογή του κατάλληλου εξοπλισμού σε μηχανήματα, συστήματα Βάσεων Δεδομένων και εργαλείων λογισμικού.
- Εγκατάσταση επικοινωνίας μεταξύ των servers και των εργαλείων ανάλυσης
- Σχεδίαση του σχήματος της Αποθήκης Δεδομένων
- Δημιουργία της φυσικής οργάνωσης της Αποθήκης Δεδομένων, υλοποίηση των σχετικών δομών και των μεθόδων πρόσβασης στην Αποθήκη.
- Σχεδίαση και ανάπτυξη των προγραμμάτων που εκτελούν τη μεταφορά δεδομένων.
- Εγκατάσταση των μεταφορέων και σύνδεση με τις πηγές δεδομένων.
- Δημιουργία της Βάσης των Μετα-δεδομένων.
- Ολοκλήρωση των εφαρμογών ανάλυσης.

2.3 Μεταφορά δεδομένων από τις πηγές στην Αποθήκη Δεδομένων

Βασικός παράγοντας για την επιτυχία των Αποθηκών Δεδομένων είναι η ορθή τροφοδοσία της Αποθήκης Δεδομένων από τις πηγές. Η διαδικασία μεταφοράς δεδομένων από τις πηγές στην Αποθήκη δεδομένων είναι αρκετά πολύπλοκη καθώς πολλά προβλήματα πρέπει να αντιμετωπισθούν. Τα βήματα που ακολουθούνται κατά τη μεταφορά των δεδομένων παρουσιάζονται στο Σχήμα 2.2



Σχήμα 2.2 : Διαδικασία μεταφοράς δεδομένων

2.3.1 Εξαγωγή και Μετατροπή Δεδομένων

Η Εξαγωγή και η Μετατροπή δεδομένων εκτελούνται από τους Μεταφορείς / Μετατροπείς του Σχήματος 2.1. Για κάθε πηγή που χρησιμοποιούμε στο σύστημα εγκαθιστούμε λογισμικό που αντλεί τα δεδομένα από την πηγή, τα “καθαρίζει”, κρατώντας μόνο αυτά που είναι πραγματικά χρήσιμα και τα μετασχηματίζει με βάση ένα καθορισμένο πρότυπο. Οι μετατροπές που γίνονται στα δεδομένα αφορούν τόσο τη δομή όσο και την τιμή τους. Για παράδειγμα, το πεδίο “Ημερομηνία” ενός πίνακα μπορεί να μετασχηματιστεί στα πεδία “Χρόνος”, “Μήνας” και “Ημέρα”, ενώ οι τιμές του πεδίου “Χαρακτηρισμός” είναι πιθανόν να μετατραπούν από “Α”, “Β” κλπ σε “1”, “2” κλπ αντίστοιχα. Αυτό το λογισμικό υλοποιείται με βάση τα ιδιαίτερα χαρακτηριστικά κάθε πηγής και εγκαθίσταται σε υπολογιστές με άμεση πρόσβαση στα δεδομένα της πηγής. Οι Αποθήκες Δεδομένων χρησιμοποιούν ποικίλα εργαλεία για εξαγωγή. Η εξαγωγή δεδομένων από τις απομακρυσμένες πηγές συχνά υλοποιείται μέσω πυλών (gateways) και καθιερωμένων προτύπων διασύνδεσης εφαρμογών (όπως ODBC, Oracle Open Connect, Information Builders EDA/SQL κλπ).

Εξωτερικά εργαλεία που εγκαθίστανται για κάθε διαφορετική πηγή δεδομένων αναλαμβάνουν την εξαγωγή των δεδομένων από τις πηγές. Παράλληλα εκτελούν και μία πρώτη επεξεργασία των δεδομένων αυτών. Καθώς οι Αποθήκες Δεδομένων χρησιμοποιούνται για στρατηγικές αποφάσεις, επιβάλλεται να περιέχουν σωστά δεδομένα. Στις διάφορες πηγές όπου υπάρχει μεγάλος όγκος δεδομένων είναι πολύ πιθανόν να υπάρχουν λάθη ή ανωμαλίες.

Διάφορα εργαλεία βοηθούν στη διάγνωση των ανωμαλιών των δεδομένων και στη διόρθωσή τους όπου αυτό είναι εφικτό. Ως περιπτώσεις όπου ο καθαρισμός των δεδομένων είναι σημαντικός αναφέρονται: ασυνέπειες στο μήκος των πεδίων διαφορετικών πηγών, ασυνέπειες σχετικά με την περιγραφή των δεδομένων, ασυνεπείς τιμές δεδομένων, απουσίες εγγραφών και παραβίαση περιορισμών ακεραιότητας.

2.3.2 Ολοκλήρωση

Η διαδικασία της ολοκλήρωσης (*integration*) των δεδομένων είναι αρκετά πολύπλοκη και περιλαμβάνει τη δημιουργία και συντήρηση ενός καθολικού ιδεατού σχήματος των δεδομένων των πηγών. Αυτό το σχήμα περιλαμβάνει κάθε οντότητα που παρέχει δεδομένα από οποιαδήποτε πηγή της Αποθήκης Δεδομένων. Η Βάση των μετα-δεδομένων ενημερώνεται και συντηρεί το καθολικό σχήμα. Με βάση το καθολικό σχήμα, κάθε ποσότητα δεδομένων που έρχεται από τις πηγές πρέπει να μετασχηματιστεί ώστε να εισαχθεί στην Αποθήκη Δεδομένων.

Για παράδειγμα, σε ένα τηλεπικοινωνιακό οργανισμό, είναι πιθανόν δύο συστήματα να διαχειρίζονται χρεώσεις από διαφορετικούς πελάτες. Στις βάσεις και των δύο πληροφοριακών συστημάτων υπάρχει πίνακας “Χρέωση” με πιθανά διαφορετικό ορισμό στο κάθε σύστημα.. Μετά τη διαδικασία ολοκλήρωσης των πηγών στο καθολικό σχήμα, υπάρχει επίσης η οντότητα “Χρέωση”, ο ορισμός της οποίας προκύπτει μεν από τους επιμέρους ορισμούς, αλλά πιθανά δεν ακολουθεί επακριβώς κάποιον από τους δύο ή και τους δύο. Οι εγγραφές που αντιστοιχούν σε χρεώσεις που γίνονται στα δύο συστήματα θα πρέπει να τροποποιηθούν για να εισαχθούν στην Αποθήκη Δεδομένων.

2.3.3 Εισαγωγή δεδομένων

Τελευταίο στάδιο στη μεταφορά των δεδομένων από τις πηγές στην Αποθήκη Δεδομένων είναι η διαδικασία εισαγωγής. Κατά τη διάρκεια της εισαγωγής, τα δεδομένα επεξεργάζονται ώστε να ελεγχθούν οι περιορισμοί ακεραιότητας της Αποθήκης Δεδομένων και να γίνουν υπολογισμοί πάνω στα δεδομένα όπως αθροιστικές πράξεις και ομαδοποιήσεις. Από το αποτέλεσμα αυτών των πράξεων θα προκύψουν τα δεδομένα που θα καταγραφούν στην Αποθήκη Δεδομένων, ενώ παράλληλα ενημερώνονται τα ευρετήρια της βάσης της Αποθήκης. Κατά τη διάρκεια της εισαγωγής των δεδομένων, πρέπει να παρέχεται η δυνατότητα στο διαχειριστή της Αποθήκης δεδομένων να παρακολουθεί και να επεμβαίνει στην όλη διαδικασία. Καθώς η διαδικασία εισαγωγής δεδομένων έχει μεγάλο υπολογιστικό κόστος και στις πηγές, αλλά και στην αποθήκη δεδομένων, η διαδικασία αυτή γίνεται μαζικά σε περιοδικά χρονικά διαστήματα όπου δεν υπάρχει φορτίο στο σύστημα.

2.3.4 Ενημέρωση

Η ενημέρωση της Αποθήκης Δεδομένων είναι η διαδικασία που μεταφέρει τις αλλαγές που συμβαίνουν στα δεδομένα των πηγών εκτελώντας αντίστοιχες αλλαγές στα δεδομένα της Αποθήκης. Η διαδικασία αυτή ακολουθεί όλα τα παραπάνω βήματα (εξαγωγή, μετατροπή, ολοκλήρωση, εισαγωγή). Υπάρχουν όμως και μερικά επιπλέον ζητήματα που προκύπτουν από τη δυνατότητα διάγνωσης των μεταβολών στις πηγές, καθώς και από τον όγκο των δεδομένων που τροποποιούνται. Συνήθως, οι Αποθήκες Δεδομένων ενημερώνονται περιοδικά. Υπάρχουν όμως και περιπτώσεις όπου εφαρμογές ανάλυσης απαιτούν άμεση πρόσβαση σε τρέχοντα δεδομένα, οπότε επιβάλλεται η άμεση ενημέρωση των Αποθηκών για κάθε μεταβολή στις πηγές. Η πολιτική ενημέρωσης καθορίζεται από το διαχειριστή της Αποθήκης Δεδομένων με βάση τις ανάγκες των εφαρμογών ανάλυσης, τη διαθεσιμότητα των πηγών και τη κατάσταση του δικτύου που συνδέει την Αποθήκη με τις πηγές.

Οι τεχνικές ενημέρωσης εξαρτώνται από τα χαρακτηριστικά των πηγών. Πολλές φορές είναι δυνατή μόνο η εξαγωγή ενός ολόκληρου αρχείου ή μία βάσης δεδομένων από μία πηγή. Σε αυτή την περίπτωση, η ενημέρωση της Αποθήκης θα ισοδυναμούσε με διαγραφή όλων των δεδομένων που σχετίζονται με τη πηγή και επανεισαγωγή των εξαχθέντων δεδομένων. Πρόκειται για μία καθόλου αποδοτική λύση, που όμως πολλές φορές είναι και η μοναδική, όταν η πηγή αδυνατεί να μας δώσει πληροφορίες για τις μεταβολές που συντελούνται σε αυτή.

Δεδομένου ότι οι Αποθήκες Δεδομένων συσσωρεύουν μεγάλη ποσότητα δεδομένων, οι εφαρμογές της παραπάνω μεθόδου ενημέρωσης καθίσταται απαγορευτική. Γι' αυτό και είναι αναγκαία η διάγνωση των μεταβολών που συμβαίνουν στις πηγές (εισαγωγές, διαγραφές και τροποποιήσεις εγγραφών), ώστε σε κάθε διαδικασία ενημέρωσης να μην γίνονται περιττές διαγραφές και εισαγωγές δεδομένων που στην πραγματικότητα παραμένουν αναλλοίωτα. Σύμφωνα με τις τεχνικές προοδευτικής ενημέρωσης και συντήρησης των δεδομένων, εισάγονται στην Αποθήκη νέες εγγραφές που προκύπτουν αποκλειστικά από αντίστοιχες εισαγωγές δεδομένων στις πηγές. Ομοίως, και οι διαγραφές και τροποποιήσεις εγγραφών προκύπτουν από αντίστοιχες πράξεις δεδομένων στις πηγές.

Για να γίνει εφικτή η εφαρμογή της προοδευτικής ενημέρωσης πρέπει οι πηγές να μας δίνουν τη δυνατότητα διάγνωσης των μεταβολών που συντελούνται στα δεδομένα τους. Στις περιπτώσεις που η πηγή είναι ένα σύγχρονο σύστημα βάσης δεδομένων, υπάρχουν τρεις βασικές τεχνικές με τις οποίες είναι εφικτή η διάγνωση των μεταβολών αυτών:

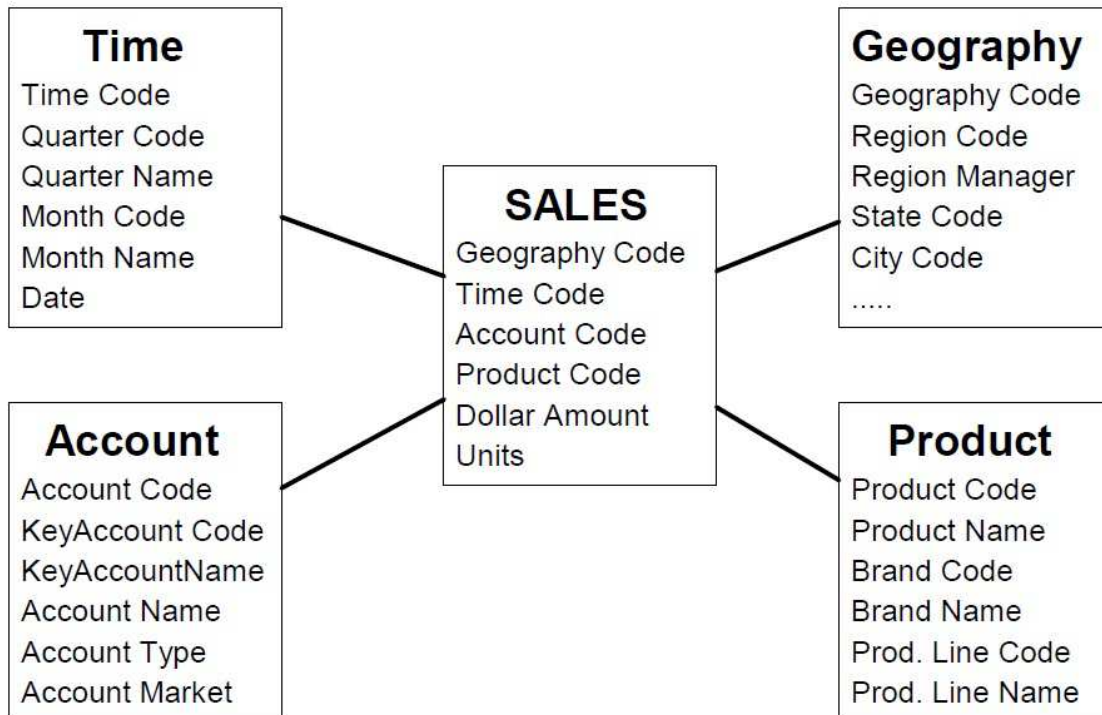
- Στιγμιότυπα: Αρκετά συστήματα βάσεων δεδομένων είναι σε θέση να εξάγουν, όταν τους ζητηθεί, στιγμιότυπα (snapshots) από πίνακες της βάσης τους. Από τα στιγμιότυπα αυτά με διάφορες μεθόδους αποδοτικής σύγκρισης μπορούμε να διαγνώσουμε τις τροποποιήσεις που συνέβηκαν στην πηγή και να ενημερωθεί σχετικά η Αποθήκη.
- Μηχανισμός καταγραφής (log): Τα περισσότερα σύγχρονα συστήματα βάσεων δεδομένων καταγράφουν όλες τις μεταβολές των δεδομένων τους και τις πράξεις που τις προκαλούν, ώστε να μπορούν να παρέχουν ομαλή εκτέλεση των δοσοληψιών. Οι μεταφορείς που εξάγουν τα δεδομένα από τέτοιες πηγές μπορούν να έχουν άμεση πρόσβαση στις μεταβολές που συντελούνται, αν τους δοθεί η δυνατότητα πρόσβασης στο αρχείο (log file) που καταγράφονται αυτές οι μεταβολές.
- Triggers: Σε περίπτωση που μία πηγή είναι ένα μοντέρνο σύστημα που παρέχει τη δυνατότητα δημιουργίας triggers, μπορούμε για κάθε πίνακα της πηγής να δημιουργήσουμε έναν trigger που θα μας ενημερώνει για οποιαδήποτε μεταβολή συμβαίνει στον πίνακα αυτόν.

2.4 Σχεδίαση Αποθηκών Δεδομένων

Καθώς οι Αποθήκες Δεδομένων χρησιμοποιούνται αποκλειστικά για την απάντηση των ερωτήσεων των εφαρμογών ανάλυσης, η σχεδίαση και η οργάνωση των δεδομένων είναι διαφορετική από τις κλασικές βάσεις δεδομένων. Τα διαγράμματα Οντοτήτων - Συσχετίσεων και οι τεχνικές κανονικοποίησης είναι οι κλασικές μέθοδοι για τη σχεδίαση των βάσεων δεδομένων των συστημάτων επεξεργασίας δοσοληψιών (OLTP). Αυτές οι μέθοδοι αποδεικνύονται συχνά ακατάλληλες για τη σχεδίαση των Αποθηκών Δεδομένων, καθώς ο στόχος τους είναι να αντιμετωπίσουν προβλήματα όπως ο πλεονασμός (redundacy), ή η ανανέωση των δεδομένων. Επιπλέον, αυτές οι μέθοδοι οδηγούν στη δημιουργία πολλών πινάκων με μικρό αριθμό πεδίων, σχήμα που έχει σαν αποτέλεσμα την εκτέλεση μεγάλου αριθμού από πράξεις JOIN, στην

περίπτωση που θέλουμε να αντλήσουμε μεγάλο όγκο αναλυτικών πληροφοριών.

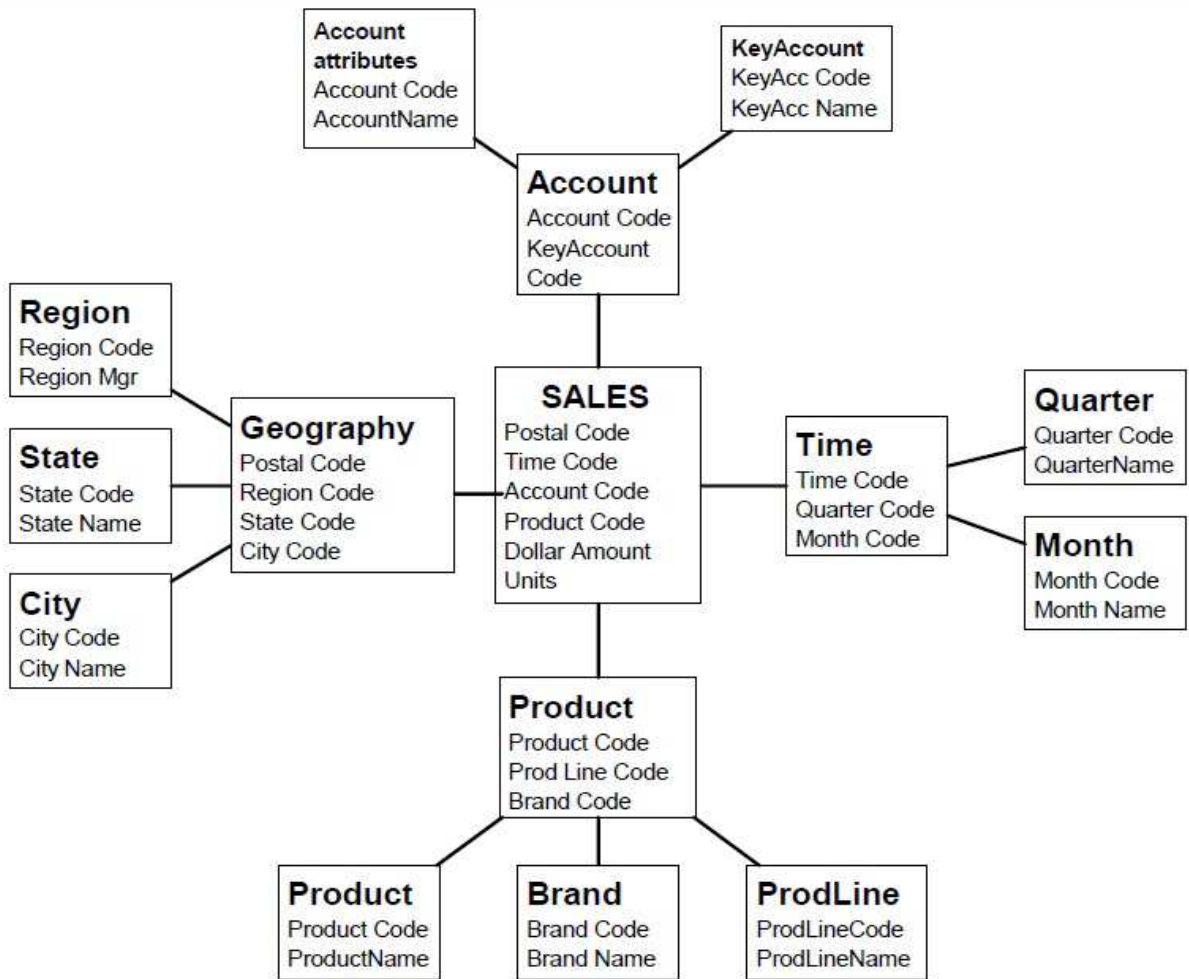
Οι πιο κατάλληλες τεχνικές για τη σχεδίαση των βάσεων των Αποθηκών Δεδομένων είναι τα *αστεροειδή σχήματα (star schemata)* και τα *σχήματα χιονονιφάδας (snowflake schemata)*. Το αστεροειδές σχήμα είναι πιο κοντά στον πολυδιάστατο χαρακτήρα των δεδομένων. Σε μία αστεροειδή βάση, υπάρχει ένας βασικός πίνακας που χαρακτηρίζεται *πίνακας συμβάντων (fact table)*. Υπάρχει επίσης ένας πίνακας για κάθε μία διάσταση. Κάθε εγγραφή του πίνακα συμβάντων αποτελείται από ένα δείκτη (ξένο κλειδί) σε μία εγγραφή κάθε ενός από τους πίνακες διαστάσεων. Κάθε *πίνακας διάστασης (dimension table)* περιλαμβάνει εγγραφές που αντιστοιχούν σε τιμές των διαστάσεων.



Σχήμα 2.3 : Παράδειγμα αστεροειδούς σχήματος

Στο Σχήμα 2.3 παρουσιάζεται ένα παράδειγμα αστεροειδούς σχήματος. Πρόκειται για το σχήμα Αποθήκης Δεδομένων που περιλαμβάνει δεδομένα σχετικά με τις πωλήσεις προϊόντων σε διάφορες πόλεις. Ο πίνακας των πωλήσεων (SALES) είναι στην προκειμένη περίπτωση ο πίνακας συμβάντων. Παρατηρούμε -για παράδειγμα- ότι μπορεί να εκτελεστεί οποιαδήποτε ερώτηση που συσχετίζει πωλήσεις, με τους λογαριασμούς (accounts) που παράγονται τα προϊόντα και με τις γεωγραφικές περιοχές που αυτά πωλούνται, εκτελώντας μόνο δύο πράξεις JOIN.

Η κύρια αδυναμία των αστεροειδών σχημάτων εντοπίζεται στον τρόπο με τον οποίο εκφράζουν τις ιεραρχίες των διαστάσεων. Για παράδειγμα, στη διάσταση χρόνος υπάρχει μία προφανής ιεραρχία μεταξύ ημερών, μηνών, ετών, κλπ. Μια εναλλακτική μοντελοποίηση των ιεραρχιών γίνεται από τα σχήματα χιονονιφάδας. Στο Σχήμα 2.4 παρουσιάζεται το παράδειγμα βάσης ίδιου περιεχομένου με τη βάση του Σχήματος 2.3, αλλά οργανωμένη σύμφωνα με το σχήμα χιονονιφάδας. Πρόκειται για μία βελτίωση του αστεροειδούς σχήματος, όπου η ιεραρχία των διαστάσεων αναπαρίσταται κανονικοποιώντας τους πίνακες των διαστάσεων.



Σχήμα 2.4 Παράδειγμα σχήματος χιονονιφάδας.

Σε περιπτώσεις σχεδίασης Αποθηκών Δεδομένων με δεδομένα πολύπλοκης δομής είναι πιθανό, περισσότεροι του ενός πίνακες συμβάντων να έχουν κοινούς πίνακες διαστάσεων. Για παράδειγμα, οι παραγγελίες και οι πωλήσεις έχουν κοινές τις περισσότερες διαστάσεις.

Εκτός από τους πίνακες συμβάντων και διαστάσεων, είναι πιθανόν να υπάρχουν και επιπρόσθετοι πίνακες με συγκεντρωτικά, προ-υπολογισμένα δεδομένα στην Αποθήκη Δεδομένων. Στην πιο απλή περίπτωση, τα συγκεντρωτικά δεδομένα αντιστοιχούν στην ομαδοποίηση των εγγράφων των πινάκων συμβάντων στη βάση συνδυασμού διαστάσεων. Στη βάση του παραδείγματος των σχημάτων 2.3 και 2.4 μπορεί να προστεθούν πίνακες οι οποίοι να περιέχουν τις συνολικές πωλήσεις προϊόντων ανά γεωγραφική περιοχή και μονάδα χρόνου. Στην πραγματικότητα δηλαδή, πρόκειται για πίνακες που περιέχουν πληροφορίες που προκύπτουν από τα δεδομένα του πίνακα συμβάντων. Η σκοπιμότητα ύπαρξης αυτών των πινάκων κρίνεται από την άφιξη σχετικών ερωτήσεων στη βάση από τις εφαρμογές ανάλυσης. Εναλλακτική της δημιουργίας τέτοιων πινάκων, είναι η εισαγωγή στους πίνακες συμβάντων εγγράφων που θα περιέχουν συγκεντρωτικές πληροφορίες για μερικές από τις διαστάσεις. Στην περίπτωση που θέλουμε να εισάγουμε στον πίνακα των πωλήσεων (SALES) συγκεντρωτικές τιμές ανά περιοχή και μονάδα χρόνου, θα εισάγουμε εγγραφές που

- τα πεδία που αντιστοιχούν στις διαστάσεις θα έχουν τις αντίστοιχες τιμές,
- τα πεδία που αντιστοιχούν στη τιμή που μετρά κάθε εγγραφή (units) θα υπάρχει η συγκεντρωτική τιμή, και
- στα πεδία των διαστάσεων που αθροίζονται (που δεν μας ενδιαφέρουν πλέον, δηλαδή) θα υπάρχουν τιμές NULL.

Η παρακάτω εγγραφή αντιστοιχεί στις συνολικές πωλήσεις του Ιουνίου στην περιοχή με κωδικό 16232 που παρέχει η αποθήκη δεδομένων του σχήματος 2.3 και 2.4.

Postal code	Time Code	Account Code	Product Code	Dollar Amount	Units
16232	Ιούλιος 1997	<i>null</i>	<i>null</i>	1654000	6520

2.5 Ειδικά θέματα των συστημάτων Αποθηκών Δεδομένων

Οι Αποθήκες Δεδομένων συνήθως περιέχουν εξαιρετικά μεγάλες ποσότητες δεδομένων. Η αποδοτική απάντηση των ερωτήσεων απαιτεί ευφυείς μεθόδους πρόσβασης και τεχνικές επεξεργασίας των ερωτήσεων. Οι συνήθεις λύσεις που δίνονται σχετίζονται με την ευρεία χρήση *ευρετηρίων (indexes)*. Η επιλογή των κατάλληλων ευρετηρίων που θα δημιουργηθούν είναι πολύ σημαντικό πρόβλημα της σχεδίασης της Αποθήκης Δεδομένων. Το επόμενο βήμα αφορά τη σωστή διαχείριση των παραπάνω δομών. Η βελτιστοποίηση των ερωτήσεων είναι επίσης ένα σημαντικό ζήτημα. Καθώς αρκετές από τις ερωτήσεις που θέτονται στο σύστημα δεν είναι αποδοτικό να απαντηθούν με τη χρήση των ευρετηρίων, είναι αναγκαία η βελτιστοποίηση ακόμα και των μεθόδων που εκτελούν σειριακή αναζήτηση. Οι δυνατότητες που παρέχουν τα παράλληλα συστήματα αποδεικνύονται συχνά αποτελεσματικές.

2.5.1 Δομές Ευρετηρίων και η χρήση τους

Υπάρχουν πολλές τεχνικές επεξεργασίας ερωτήσεων που εκμεταλλεύονται αποδοτικά τα ευρετήρια. Για παράδειγμα ερωτήσεις με πολλαπλές συνθήκες μπορούν να απαντηθούν με τη χρήση της τομής και ένωσης των δεδομένων ευρετηρίων. Αυτές οι πράξεις μπορούν να χρησιμοποιηθούν για σημαντική μείωση του κόστους απάντησης των ερωτήσεων, ενώ συχνά παρακάμπτεται η πρόσβαση στους πίνακες με τα δεδομένα.

Τα συστήματα Αποθηκών Δεδομένων χρησιμοποιούν bitmap ευρετήρια, που υποστηρίζουν αποδοτικά πράξεις όπως ένωση, ή τομή. Έστω μια σελίδα σε φύλλο ενός ευρετηρίου που αντιστοιχεί στην τιμή A. Μία τέτοια σελίδα περιέχει μία λίστα από διευθύνσεις εγγραφών που περιέχουν τη τιμή A. Τα bitmap ευρετήρια δομούν τη λίστα των διευθύνσεων ως ένα διάνυσμα από δυαδικές τιμές (0,1), που έχει μία δυαδική μεταβλητή (bit) για κάθε εγγραφή. Η μεταβλητή αυτή παίρνει τιμή 1 αν η εγγραφή στην οποία αντιστοιχεί περιέχει τη τιμή A. Η αποδοχή των bitmap ευρετηρίων στηρίζεται στο γεγονός ότι οι αναπαράσταση της λίστας των διευθύνσεων των εγγραφών σε διάνυσμα από bits επιταχύνει πράξεις όπως σύνδεση, τομή, ένωση και ομαδοποίηση, καθώς αυτές μετατρέπονται σε λογικές πράξεις πάνω σε πίνακες από bits και εκτελούνται γρήγορα.

Εκτός από τα ευρετήρια τιμών σε ένα πίνακα, η δομή των αστεροειδών σχημάτων επιβάλλει την χρήση των *ευρετηρίων σύνδεσης (join indices)*. Τα ευρετήρια αυτού του είδους παρέχουν τη συσχέτιση τη τιμής ενός ξένου κλειδιού ενός πίνακα με την αντίστοιχη τιμή του κλειδιού του πίνακα στον οποίο αναφέρεται. Σε μία βάση με αστεροειδές σχήμα μπορούμε να συσχετίσουμε τον πίνακα συμβάντων με τους πίνακες των διαστάσεων με τη χρήση των ευρετηρίων σύνδεσης. Για παράδειγμα, στη βάση του Σχήματος 2.3 μπορεί να υπάρχει ένα ευρετήριο σύνδεσης στον κωδικό Postal Code που κρατά για κάθε διαφορετική πόλη τις εγγραφές στον πίνακα των συμβάντων που αντιστοιχούν στην πόλη αυτή.

2.5.2 Μετατροπή Πολύπλοκων Ερωτήσεων

Η εύρεση κατάλληλου μετασχηματισμού των ερωτήσεων ώστε να απαντιούνται αποδοτικά αποδεικνύεται επίσης αρκετά σημαντική. Στη περιοχή των Αποθηκών Δεδομένων συναντάμε κλασικά θέματα, όπως αυτό της επεξεργασίας των φωλιασμένων ερωτήσεων. Οι ερωτήσεις που περιέχουν φωλιασμένες υποερωτήσεις καταναλώνουν γενικά πολύ χρόνο για να απαντηθούν. Υπάρχουν αρκετές τεχνικές που μετασχηματίζουν τις φωλιασμένες και ερωτήσεις πολλών συνθηκών.

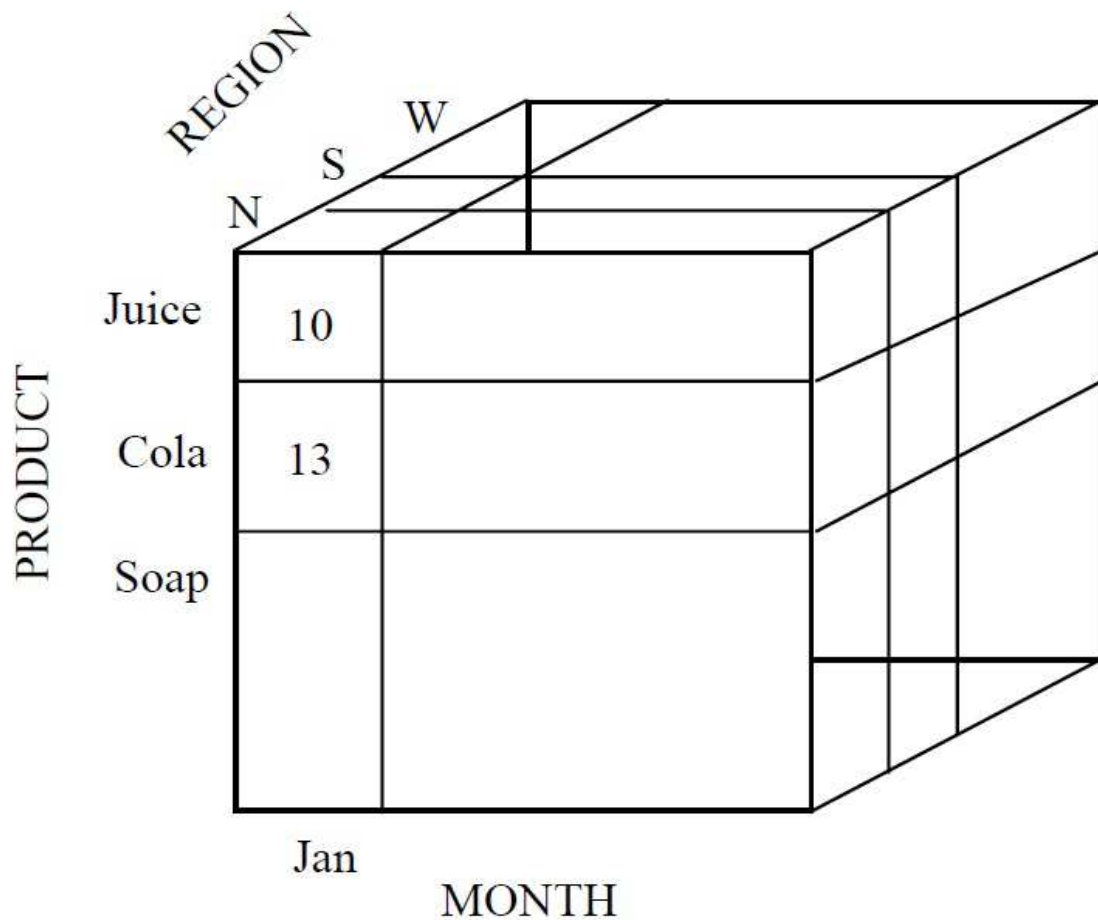
Η αποδοτική εκτέλεση των ερωτήσεων που περιλαμβάνουν πράξεις join μεταξύ πινάκων είναι επίσης αναγκαία. Ειδικές λύσεις προτείνονται για βάσεις με αστεροειδή σχήματα, ή σχήματα χιονοφάδας, καθώς η εκτέλεση ερωτήσεων στην Αποθήκη δεδομένων περιλαμβάνει, σε όλες σχεδόν τις περιπτώσεις, join μεταξύ του πίνακα συμβάντων και πινάκων διαστάσεων. Σε αρκετές περιπτώσεις εκτέλεσης join μεταξύ του πίνακα συμβάντων και περισσότερων του ενός πινάκων διαστάσεων, ακολουθείται η παρακάτω στρατηγική. Η αποθήκη εκτελεί ερωτήσεις και υπολογίζει το πλήρες καρτεσιανό γινόμενο μεταξύ των πινάκων των διαστάσεων (έχοντας πιθανά περιορίσει το εύρος των τιμών που λαμβάνουν μέρος με βάση τη συνθήκη επιλογής της ερώτησης). Κατόπιν εκτελεί μία απλή πράξη join μεταξύ του καρτεσιανού γινομένου και του πίνακα συμβάντων. Η παραπάνω μέθοδος χρησιμοποιείται για να αποφευχθεί η εκτέλεση πολλαπλών join στα οποία θα συμμετέχει ο πίνακας συμβάντων, ο οποίος συνήθως περιέχει συγκριτικά με τους πίνακες διαστάσεων, πολλαπλάσιο αριθμό εγγραφών.

2.6 Μοντέλα συστημάτων αναλυτικής επεξεργασίας

Η αναλυτική επεξεργασία δεδομένων είναι τμήμα των εφαρμογών στήριξης αποφάσεων και των στρατηγικών πληροφοριακών συστημάτων. Η λειτουργία της αναλυτικής επεξεργασίας χαρακτηρίζεται από δυναμική πολυδιάστατη ανάλυση των δεδομένων του οργανισμού. Οι εφαρμογές που εκτελούν συνεχή αναλυτική επεξεργασία (OLAP), χαρακτηρίζονται από τις συνεχείς ερωτήσεις που εκτελούν πάνω στα δεδομένα του οργανισμού. Οι ερωτήσεις αυτές έχουν συγκεκριμένη και πολύπλοκη δομή, ενώ η πληροφορία που αντλούν έχει πολυδιάστατο χαρακτήρα. Για παράδειγμα, μια εφαρμογή που εκτελεί αναλυτική επεξεργασία στα δεδομένα των πωλήσεων ενός οργανισμού, συνεχώς εκτελεί ερωτήσεις για να μπορεί να έχει συγκεντρωτικά δεδομένα για τις πωλήσεις ανά προϊόν, ανά μήνα και ανά περιοχή. Η παρουσίαση των αποτελεσμάτων των πωλήσεων μπορεί να προκαλέσει τον χρήστη να εκτελέσει μία πιο συγκεντρωτική ερώτηση, ώστε να πάρει δεδομένα των ετήσιων πωλήσεων ανά προϊόν και περιοχή, ή να εκτελέσει μια πιο λεπτομερή ερώτηση παίρνοντας τις μηνιαίες πωλήσεις κάθε προϊόντος ανά συγκεκριμένο πελάτη.

2.6.1 Πολυδιάστατα Μοντέλα Δεδομένων

Οι πίνακες των σχεσιακών βάσεων δεδομένων περιέχουν εγγραφές οι οποίες αποτελούνται από πεδία. Σε φυσιολογικές σχεσιακές βάσεις δεδομένων, ένα υποσύνολο των πεδίων ενός πίνακα συνθέτουν το κλειδί του. Αντίθετα τα πολυδιάστατα μοντέλα δεδομένων περιέχουν n-διάστατους πίνακες που συχνά αποκαλούνται *υπερκύβοι* (*cubes* ή *hypercubes*). Κάθε διάσταση έχει μία ιεραρχία επιπέδων. Για παράδειγμα, η διάσταση "Γεωγραφική τοποθεσία" έχει τα επίπεδα πόλη, περιοχή, χώρα. Οι τιμές (μετρικές) που περιέχουν οι υπερκύβοι αντιστοιχούν στις στήλες των σχεσιακών πινάκων.



Σχήμα 2.5 Παράδειγμα υπερκύβου

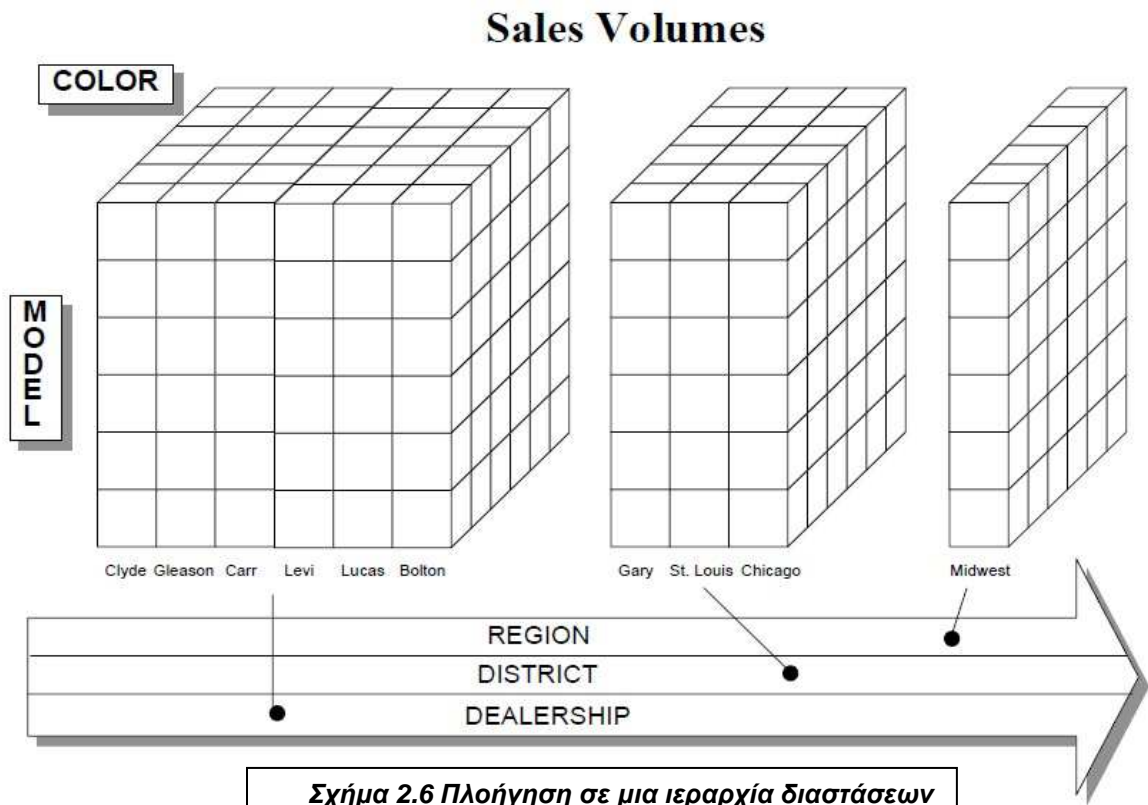
Στο Σχήμα 2.5 παρουσιάζεται το μοντέλο των δεδομένων ενός υπερκύβου που παρέχει δεδομένα για τις πωλήσεις των προϊόντων. Σύμφωνα με το παράδειγμα οι πωλήσεις του προϊόντος “Cola” το μήνα Ιανουάριο στην βόρεια περιοχή ήταν 13. Οι τιμές “Cola”, “JAN” και “N” στο συγκεκριμένο παράδειγμα είναι οι τιμές των διαστάσεων “PRODUCT”, “MONTH” και “REGION” αντίστοιχα, ενώ το 13 αποτελεί τιμή των δεδομένων του υπερκύβου. Είναι πιθανόν η τιμή σε μία θέση του υπερκύβου να περιέχει συγκεντρωτική τιμή μίας μετρικής. Στο συγκεκριμένο παράδειγμα, ο αριθμός 13 αφορά τις πωλήσεις όλων των ημερών του μήνα, σε όλες τις πόλεις στην Βόρεια περιοχή. Προφανώς, αυτός ο αριθμός έχει προκύψει από αναλυτικότερα δεδομένα τα οποία μπορεί και να υπάρχουν σε άλλον υπερκύβο. Οι διαστάσεις του υπερκύβου λειτουργούν ως δείκτες στα δεδομένα του. Σε κάθε ερώτηση μπορούμε να έχουμε πρόσβαση στις τιμές των δεδομένων του κύβου με τη χρήση τιμών των διαστάσεων. Με τη χρήση τιμών για όλες τις διαστάσεις παίρνουμε μία απλή τιμή από τον υπερκύβο. Μπορούμε να χρησιμοποιήσουμε τιμές για μερικές από τις διαστάσεις του υπερκύβου. Αν κάνουμε μια ερώτηση χρησιμοποιώντας τιμές έστω από δύο διαστάσεις του κύβου, τότε θα πάρουμε συγκεντρωτική πληροφορία για αυτές τις δύο διαστάσεις. Αν στο παράδειγμα του Σχήματος 2.5 εκτελέσουμε μία αναζήτηση στα δεδομένα του κύβου με βάση μόνο συγκεκριμένες τιμές για δυο διαστάσεις (PRODUCT = “Cola” και MONTH = “JAN”) τότε θα πάρουμε το άθροισμα των πωλήσεων του προϊόντος “Cola” το μήνα Ιανουάριο σε όλες τις περιοχές. Αν εκτελέσουμε μία αναζήτηση στα δεδομένα του κύβου για την τιμή PRODUCT = “Cola” θα πάρουμε τις συνολικές πωλήσεις του συγκεκριμένου προϊόντος. Επίσης, μας δίνεται η δυνατότητα αναζητήσεων χωρίς να δίνουμε συγκεκριμένες τιμές στις διαστάσεις ή να δίνουμε περιοχή τιμών. Μπορούμε, για παράδειγμα, να πάρουμε τις συγκεντρωτικές πωλήσεις όλων των προϊόντων ανά προϊόν και μήνα για τους μήνες Ιανουάριος έως Μάρτιο.

2.6.2 Πράξεις στους υπερκύβους

Οι υπερκύβοι μας δίνουν τη δυνατότητα πλοήγησης στις ιεραρχίες των διαστάσεών τους. Η πλοήγηση είναι δυνατή από τις πράξεις οι οποίες μας παρέχονται. Οι πράξεις που συνήθως γίνονται στους υπερκύβους είναι οι παρακάτω:

Roll-up: Πρόκειται για πράξη με την οποία εκτελούμε ένα βήμα ανόδου στην ιεραρχία μιας διάστασης. Στο παράδειγμα του Σχήματος 2.6, έχουμε αρχικά ένα κύβο που αποτελείται από τρεις διαστάσεις: Χρώμα (*Color*), Μοντέλο (*Model*) και Γεωγραφία (*Geography*). Η διάσταση Γεωγραφία έχει τρία επίπεδα: κατάσταση (*dealership*), περιοχή (*district*) και περιφέρεια (*region*). Μία πράξη roll-up στη διάσταση Γεωγραφία θα μας έδινε έναν νέο κύβο που θα περιείχε αθροιστικές πωλήσεις προϊόντων ανά περιοχή, χρώμα και μοντέλο. Ο κύβος που προκύπτει από την πράξη περιέχει πιο ομαδοποιημένα δεδομένα, με βάση τη διάσταση στην οποία έγινε η ομαδοποίηση. Η ανάβαση στην ιεραρχία μπορεί να συνεχιστεί με όμοιο τρόπο.

Drill-down: Είναι η αντίστροφη πράξη του roll-up, όπου πάμε από ένα υψηλότερο επίπεδο ιεραρχίας μίας διάστασης σε ένα χαμηλότερο. Στον πίνακα του Σχήματος 2.6, μία πράξη drill-down στη διάσταση Γεωγραφία, από το επίπεδο περιφέρειας, στον τελευταίο κύβο, στο επίπεδο καταστήματος, θα μας έδινε τον αρχικό κύβο.



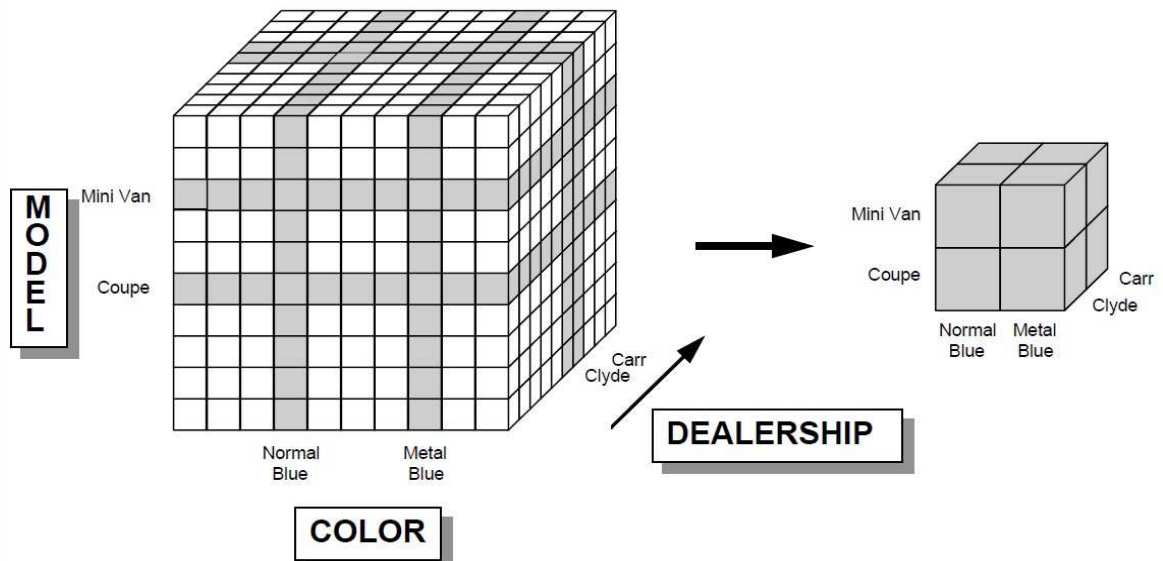
Slicing: Πρόκειται για πράξη επιλογής δεδομένων σε μία συγκεκριμένη διάσταση. Ένα επίπεδο (*slice*) είναι ένα υποσύνολο ενός υπερκύβου σύμφωνα με μία περιοχή τιμών ή μια συγκεκριμένη τιμή ενός επιπέδου διάστασης. Στο παράδειγμα του σχήματος 2.7 κάνουμε τις εξής επιλογές:

στην διάσταση Μοντέλου κρατάμε μόνο τις τιμές SPORTS COUPE και MINI VAN.

στην διάσταση Γεωγραφίας κρατάμε μόνο τα καταστήματα CARR και CLYDE.

στην διάσταση Χρώματος κρατάμε μόνο τις τιμές METAL BLUE και NORMAL BLUE

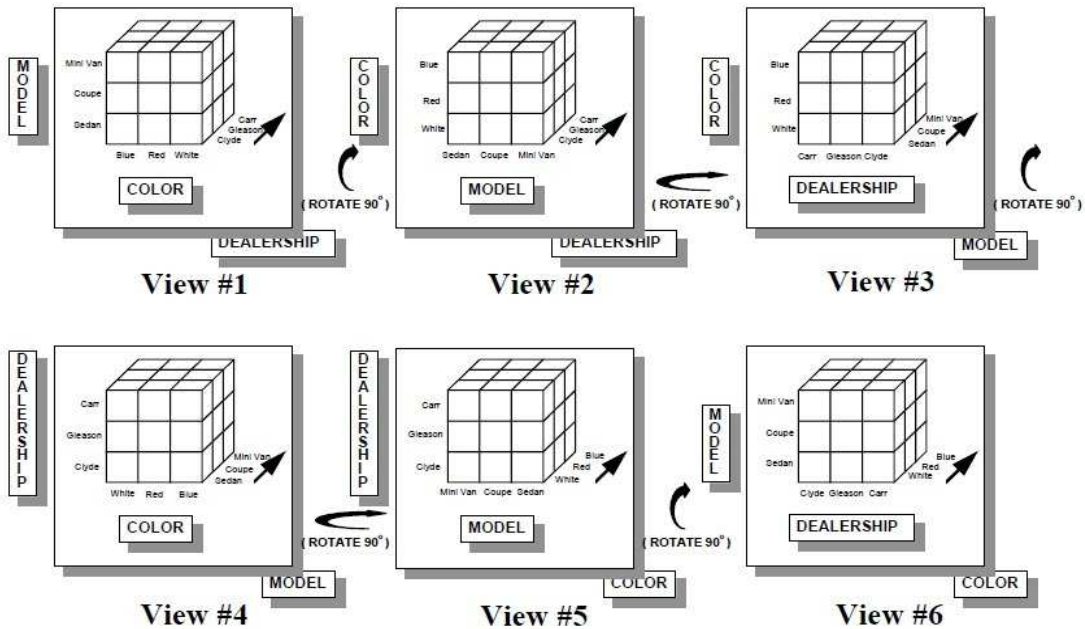
Sales Volumes



Σχήμα 2.7 Slicing

Pivoting: Πρόκειται για πράξη αλλαγής της διάταξης των διαστάσεων ώστε να διευκολυνθεί η ανάλυση. Κατά το pivoting, δεν μεταβάλλονται ούτε μειώνονται τα δεδομένα του υπερκύβου. Απλά αλλάζει ο τρόπος παρουσίασής τους στην εφαρμογή ανάλυσης. Στο σχήμα 2.8 φαίνονται οι διαφορετικοί τρόποι παρουσίασης ενός κύβου.

Sales Volumes



Σχήμα 2.8 Pivoting

3 Συστήματα υποστήριξης αποφάσεων και OLAP

Οι επιχειρήσεις άρχισαν να εκμεταλλεύονται τα online δεδομένα για να παίρνουν καλύτερες αποφάσεις για τις δραστηριότητάς τους, όπως ποια στοιχεία να κρατούν στην αποθήκη και ποιος είναι ο καλύτερος τρόπος να απευθύνονται στους πελάτες για να αυξήσουν τις πωλήσεις τους. Πολλά από τα ερωτήματά τους είναι μάλλον περίπλοκα και κάποιοι τύποι πληροφοριών δεν μπορούν να εξαχθούν ακόμα και με τη χρήση SQL. Είναι διαθέσιμες διάφορες τεχνικές και εργαλεία για να βοηθήσουν στην υποστήριξη αποφάσεων. Διάφορα εργαλεία ανάλυσης δεδομένων επιτρέπουν στους αναλυτές να βλέπουν τα δεδομένα με διαφορετικούς τρόπους. Άλλα εργαλεία ανάλυσης προϋπολογίζουν συνοπτικές πληροφορίες πολύ μεγάλων ποσοτήτων δεδομένων, για να δώσουν γρήγορη απόκριση σε ερωτήματα. Η τυποποίηση SQL:1999 περιέχει τώρα επιπλέον δομές για υποστήριξη ανάλυσης των δεδομένων. Μια άλλη προσέγγιση για να πάρουμε γνώσεις από τα δεδομένα είναι να χρησιμοποιήσουμε *εξόρυξη δεδομένων (data mining)*, που στοχεύει να εντοπίσει διάφορα μοτίβα σε μεγάλο όγκο δεδομένων. Η εξόρυξη δεδομένων συμπληρώνει διάφορους τύπους στατιστικών τεχνικών με παρόμοιους στόχους.

Τα δεδομένα κειμένου επίσης, έχουν αναπτυχθεί εκρηκτικά. Τα δεδομένα κειμένου είναι μη δομημένα, αντίθετα με τα δομημένα δεδομένα των σχεσιακών βάσεων δεδομένων. Οι ερωτήσεις σε μη δομημένα δεδομένα κειμένου αναφέρονται ως *ανάκληση πληροφοριών*. Τα συστήματα ανάκλησης πληροφοριών έχουν πολλά κοινά με τις βάσεις δεδομένων, ιδιαίτερα στην αποθήκευση και ανάκληση δεδομένων σε δευτερεύοντες χώρους αποθήκευσης. Ωστόσο, η έμφαση στα πληροφοριακά συστήματα είναι διαφορετική από αυτή στις βάσεις δεδομένων, που επικεντρώνονται σε θέματα, όπως σε ερωτήσεις βασισμένες σε λέξεις κλειδιά, στην σχετικότητα των εγγράφων του ερωτήματος και στην ανάλυση, κατηγοριοποίηση και ευρετήριο των εγγράφων.

Αυτό το κεφάλαιο καλύπτει υποστήριξη αποφάσεων, όπως online αναλυτική επεξεργασία και εξόρυξη δεδομένων και ανάκληση πληροφοριών

3.1 Συστήματα Υποστήριξης Αποφάσεων

Οι εφαρμογές βάσεων δεδομένων μπορούν να κατηγοριοποιηθούν σε εφαρμογές επεξεργασίας συναλλαγών και υποστήριξης αποφάσεων. Τα συστήματα υποστήριξης αποφάσεων χρησιμοποιούνται ευρέως σήμερα και οι εταιρείες έχουν μαζέψει μια μεγάλη ποσότητα πληροφοριών που έχουν δημιουργηθεί από αυτά τα συστήματα.

Για παράδειγμα, οι βάσεις δεδομένων εταιρειών συνήθως περιέχουν μεγάλες ποσότητες πληροφοριών για πελάτες και συναλλαγές. Το μέγεθος της αποθήκευσης των πληροφοριών που απαιτείται μπορεί να φτάνει μέχρι εκατοντάδες gigabyte ή ακόμα και terabyte, για μεγάλες αλυσίδες πωλήσεων. Οι πληροφορίες συναλλαγών για κάποιον που κάνει λιανικές πωλήσεις μπορεί να περιλαμβάνουν το όνομα ή το αναγνωριστικό (όπως τον αριθμό πιστωτικής κάρτας) του πελάτη, τα πράγματα που αγόρασε, την τιμή που πλήρωσε και τις ημερομηνίες στις οποίες έγιναν οι αγορές. Οι πληροφορίες για προϊόντα που αγοράστηκαν μπορούν να περιλαμβάνουν το όνομα του προϊόντος, τον κατασκευαστή, τον αριθμό μοντέλου, το χρώμα και το μέγεθος. Οι πληροφορίες πελατών μπορεί να περιλαμβάνουν πιστωτικό ιστορικό, ετήσιο εισόδημα, διεύθυνση κατοικίας, ηλικία και ακόμα και εκπαιδευτικές γνώσεις.

Τέτοιες μεγάλες βάσεις δεδομένων μπορεί να έχουν θησαυρούς με πληροφορίες για να βοηθήσουν τις επιχειρήσεις να παίρνουν αποφάσεις, όπως ποια προϊόντα να έχουν στην αποθήκη και ποιες εκπτώσεις να προσφέρουν. Για παράδειγμα, μια εταιρεία λιανικής μπορεί να παρατηρήσει μια ξαφνική αύξηση στις αγορές των φανελένιων υποκαμίσων σε κάποια περιοχή και να συνειδητοποιήσει ότι μπορεί να υπάρχει μια τέτοια μόδα και να αρχίσει να βάζει σε στοκ ένα μεγάλο αριθμό από τέτοια πουκάμισα σε καταστήματα της περιοχής. Ένα άλλο παράδειγμα είναι για μια εταιρεία αυτοκινήτων που μπορεί να βρει, κάνοντας ερωτήσεις στην βάση δεδομένων της, ότι τα περισσότερα από τα μικρά σπορ αυτοκίνητά της αγοράζονται από νεαρές γυναίκες των οποίων το ετήσιο εισόδημα είναι πάνω από 50.000€. Η εταιρεία μπορεί μετά να προσαρμοσθεί για να προσελκύσει περισσότερες τέτοιες γυναίκες για να αγοράσουν μικρά σπορ αυτοκίνητα και μπορεί να αποτύχει να χάσει χρήματα προσπαθώντας να προσελκύσει άλλες κατηγορίες ατόμων να αγοράσουν αυτά τα αυτοκίνητα. Και στις δυο περιπτώσεις, η εταιρεία έχει προσδιορίσει μοτίβα στην συμπεριφορά των πελατών και έχει χρησιμοποιήσει τα

μοτίβα για να πάρει επιχειρησιακές αποφάσεις.

Η αποθήκευση και η ανάκληση των δεδομένων για υποστήριξη αποφάσεων δημιουργεί διάφορα προβλήματα:

Αν και πολλά ερωτήματα υποστήριξης αποφάσεων μπορούν να γραφτούν σε SQL, άλλα είτε δεν μπορούν να εκφραστούν σε SQL είτε δεν μπορούν να εκφραστούν εύκολα σε SQL. Έχουν προταθεί συνεπώς, διάφορες επεκτάσεις της SQL για να κάνουν ευκολότερη την ανάλυση δεδομένων. Η περιοχή της *online αναλυτικής επεξεργασίας* (*online analytical processing* - OLAP) σχετίζεται με εργαλεία και τεχνικές για ανάλυση δεδομένων, που μπορεί να δώσει σχεδόν άμεσες απαντήσεις σε ερωτήματα που ζητούν συνοπτικά δεδομένα/ ακόμα και αν η βάση δεδομένων μπορεί να είναι πολύ μεγάλη. Στην Ενότητα 3.2, μελετούμε επεκτάσεις της SQL για ανάλυση δεδομένων και τεχνικές για online αναλυτική επεξεργασία.

Οι γλώσσες ερωτημάτων βάσεων δεδομένων δεν είναι κατάλληλες για την απόδοση λεπτομερών **στατιστικών αναλύσεων** των δεδομένων. Υπάρχουν διάφορα πακέτα, όπως τα SAS και S++, που βοηθούν στην στατιστική ανάλυση. Τέτοια πακέτα έχουν διασυνδεθεί με βάσεις δεδομένων, για να επιτρέψουν σε μεγάλους όγκους δεδομένων να αποθηκεύονται στην βάση δεδομένων και να ανακλώνται αποτελεσματικά για ανάλυση. Το πεδίο της στατιστικής ανάλυσης είναι ένα μεγάλο μέρος από μόνο του.

Οι τεχνικές ανακάλυψης γνώσης προσπαθούν να ανακαλύψουν αυτόματα **στατιστικούς** κανόνες και μοτίβα από δεδομένα. Το πεδίο της *εξόρυξης δεδομένων* συνδυάζει **τεχνικές** ανακάλυψης γνώσης που έχουν ανακαλυφθεί από ερευνητές της τεχνητής **νοημοσύνης** και στατιστικούς αναλυτές, με αποτελεσματικές τεχνικές χειρισμού, που τους επιτρέπουν να χρησιμοποιούνται σε πολύ μεγάλες βάσεις δεδομένων. Η Ενότητα 3.3 συζητά την εξόρυξη δεδομένων.

Μεγάλες εταιρείες έχουν διάφορες πηγές με δεδομένα που πρέπει να χρησιμοποιήσουν για να πάρουν αποφάσεις. Οι πηγές μπορεί να αποθηκεύουν τα δεδομένα σε διαφορετικά σχήματα. Για λόγους απόδοσης, (όπως επίσης και για λόγους ελέγχου της επιχείρησης) οι πηγές δεδομένων συνήθως δεν επιτρέπουν άλλα μέρη της εταιρείας να ανακαλούν δεδομένα κατ' απαίτηση.

Για να εκτελούνται αποτελεσματικά τα ερωτήματα σε τέτοια διαφορετικά δεδομένα, οι εταιρείες έχουν δημιουργήσει *αποθήκες δεδομένων*. Οι αποθήκες συλλέγουν δεδομένα από πολλές πηγές σε ένα ενοποιημένο σχήμα, σε μια θέση. Έτσι, παρέχουν στον χρήστη μια μόνο ομοιόμορφη διασύνδεση στα δεδομένα. Μελετούμε θέματα για τη δημιουργία και συντήρηση μιας αποθήκης στην Ενότητα 3.4.

Η περιοχή της **υποστήριξης αποφάσεων** μπορούμε να θεωρήσουμε ότι καλύπτει όλες τις παραπάνω περιοχές, αν και μερικά άτομα χρησιμοποιούν τον όρο με μια πιο στενή έννοια, που αποκλείει στατιστική ανάλυση και εξόρυξη δεδομένων.

3.2 Ανάλυση Δεδομένων και OLAP

Αν και η περίπλοκη στατιστική ανάλυση είναι καλύτερα να αφεθεί στα στατιστικά πακέτα, οι βάσεις δεδομένων θα πρέπει να υποστηρίζουν απλές, συνηθισμένες μορφές ανάλυσης δεδομένων. Αφού τα δεδομένα που αποθηκεύονται σε βάσεις δεδομένων είναι συνήθως μεγάλα σε όγκο, θα πρέπει να συνοψίζονται με κάποιον τρόπο αν θέλουμε να παράγουμε πληροφορίες που θα χρησιμοποιήσουμε.

Τα εργαλεία OLAP υποστηρίζουν διαλογική ανάλυση συνοπτικών πληροφοριών. Έχουν αναπτυχθεί διάφορες επεκτάσεις της SQL για να υποστηρίξουν εργαλεία OLAP. Υπάρχουν πολλές συνηθισμένες εργασίες που δεν μπορούν να γίνουν χρησιμοποιώντας τις βασικές συνοπτικές λειτουργίες SQL και τις δυνατότητες ομαδοποίησης. Παραδείγματα περιλαμβάνουν την εύρεση ποσοστών ή αθροιστικών κατανομών, ή συνοπτικές λειτουργίες σε κυλιόμενα παράθυρα, με σειριακά ταξινομημένα δεδομένα. Έχουν πρόσφατα προταθεί διάφορες επεκτάσεις της SQL που υποστηρίζουν τέτοιες εργασίες και έχουν υλοποιηθεί σε προϊόντα όπως τα Oracle και IBM DB2.

3.2.1 Online Αναλυτική Επεξεργασία

Η στατιστική ανάλυση συνήθως απαιτεί ομαδοποίηση πολλών ιδιοτήτων. Φανταστείτε μια εφαρμογή, όπου ένα κατάστημα θέλει να βρει ποια είδη υφασμάτων είναι δημοφιλή. Ας υποθέσουμε ότι τα υφάσματα χαρακτηρίζονται από το όνομά τους, χρώμα και μέγεθος και ότι έχουμε μια σχέση *sales* με το σχήμα *sales(item-name, color, size, number)*. Υποθέστε ότι το *item-name* μπορεί να πάρει τις τιμές (skirt, dress, shirt, pant) (φούστα, φόρεμα, πουκάμισο, παντελόνι, αντίστοιχα), το *color* μπορεί να πάρει τις τιμές (dark, pastel, white) (σκούρο, παστέλ, άσπρο, αντίστοιχα) και το *size* μπορεί να πάρει τις τιμές (small, medium, large) (μικρό, μεγάλο, μεσαίο, αντίστοιχα).

Αν δοθεί μια σχέση που χρησιμοποιείται για ανάλυση δεδομένων, μπορούμε να προσδιορίσουμε κάποιες από τις ιδιότητες της ως ιδιότητες **μέτρησης**, αφού μετρούν κάποια τιμή την οποία συνοψίζουν. Για παράδειγμα, η ιδιότητα *number* της σχέσης *sales* είναι μια ιδιότητα μέτρησης, που μετρά τον αριθμό των μονάδων των προϊόντων που έχουν πουληθεί. Μερικές (ή όλες) από τις άλλες ιδιότητες της σχέσης προσδιορίζονται ως **ιδιότητες μέτρησης**, αφού ορίζουν τις διαστάσεις τις οποίες μετρούν οι ιδιότητες και συνοψίζουν τις ιδιότητες μέτρησης. Στη σχέση *sales*, τα *item-name*, *color* και *size* είναι ιδιότητες διάστασης. (Μια πιο ρεαλιστική έκδοση της σχέσης *sales* θα είχε επιπλέον διαστάσεις, όπως την ώρα και την θέση της πώλησης και επιπλέον μετρήσεις, όπως την αξία σε χρήματα της πώλησης).

Τα δεδομένα μπορούν να μοντελοποιηθούν ως ιδιότητες διάστασης και οι ιδιότητες μέτρησης ονομάζονται **πολυδιάστατα δεδομένα**.

Size: ALL		Color			
	dark	pastel	white	Σύνολο	
Item-name	skirt	8	35	10	53
	dress	20	10	5	35
	shirt	14	7	28	49
	pant	20	2	5	27
	Σύνολο	62	54	48	164

Εικόνα 3.1 Πίνακας διασταύρωσης των *sales* κατά *item-name* και *color*.

Για να αναλύσει τα πολυδιάστατα δεδομένα, ένας διευθυντής μπορεί να θέλει να δει τα δεδομένα διατεταγμένα όπως στον πίνακα της Εικόνας 2.1. Ο πίνακας δείχνει συνολικούς αριθμούς για διαφορετικούς συνδυασμούς των *item-name* και *color*. Η τιμή του *size* καθορίζεται να είναι all, δείχνοντας ότι οι εμφανιζόμενες τιμές είναι συνοπτικά δεδομένα όλων των τιμών του *size*.

Ο πίνακας που φαίνεται στην Εικόνα 3.1 είναι ένα παράδειγμα ενός **πίνακα διασταύρωσης (cross-tabulation)**, που επίσης ονομάζεται **συγκεντρωτικός πίνακας (pivot-table)**. Γενικά, ένας συγκεντρωτικός πίνακας είναι ένας πίνακας όπου οι τιμές μιας ιδιότητας (ας πούμε της A) αποτελούν τις επικεφαλίδες γραμμών, ενώ οι γραμμές μιας άλλης ιδιότητας (ας πούμε της B) αποτελούν τις επικεφαλίδες στηλών και οι τιμές ενός κελιού παράγονται ως εξής. Κάθε κελί μπορεί να προσδιοριστεί από τα (a_i, b_j) , όπου το a_i είναι μια τιμή για το A και το b_j είναι μια τιμή για το B. Αν υπάρχουν περισσότερες από μια εγγραφές με κάποια τιμή (a_i, b_j) η τιμή του κελιού παράγεται από αυτή την μοναδική εγγραφή (αν υπάρχει). Για παράδειγμα, θα μπορούσε να είναι η τιμή από μια ή περισσότερες ιδιότητες της εγγραφής. Αν μπορούν να υπάρχουν πολλές εγγραφές με μια τιμή (a_i, b_j) , η τιμή του κελιού πρέπει να παραχθεί συνοπτικά από τις εγγραφές με αυτή την τιμή. Στο παράδειγμά μας, η συνοπτική λειτουργία που χρησιμοποιείται είναι το άθροισμα των τιμών της ιδιότητας *number*. Στο παράδειγμά μας, ο συγκεντρωτικός πίνακας έχει επίσης μια επιπλέον στήλη και μια επιπλέον γραμμή που περιέχει τα σύνολα των κελιών στη γραμμή / στήλη. Οι περισσότεροι πίνακες διασταύρωσης έχουν τέτοιες συνοπτικές γραμμές και στήλες.

Ένας πίνακας διασταύρωσης είναι διαφορετικός από τους σχεσιακούς πίνακες, που συνήθως αποθηκεύονται στην βάση δεδομένων, αφού ο αριθμός των στηλών εξαρτάται από τα δεδομένα. Μια αλλαγή στις τιμές των δεδομένων μπορεί να καταλήξει στην προσθήκη επιπλέον

στηλών, που δεν είναι επιθυμητό. Ωστόσο, μια προβολή πίνακα διασταύρωσης είναι επιθυμητή για εμφάνιση στους χρήστες. Είναι απλό να αναπαρασταθεί ένας πίνακας διασταύρωσης με συνοπτικές τιμές σε μια σχεσιακή μορφή, με σταθερό αριθμό στηλών. Ένας πίνακας διασταύρωσης με συνοπτικές γραμμές / στήλες μπορεί να παρασταθεί χρησιμοποιώντας μια ειδική τιμή **all** που να αντιπροσωπεύει υποσύνολα, όπως στην Εικόνα 3.2. Η τυποποίηση SQL:1999 χρησιμοποιεί την τιμή **null** στη θέση του **all**, αλλά για να αποφευχθεί σύγχυση με τις κανονικές τιμές **null** θα συνεχίσουμε να χρησιμοποιούμε το **all**.

Φανταστείτε τις εγγραφές (skirt, **all**, 53) και (dress, **all**, 35). Έχουμε πάρει αυτές τις εγγραφές απαλείφοντας μεμονωμένες εγγραφές με διαφορετικές τιμές για το *color* και αντικαθιστώντας την τιμή του *number* με μια συνοπτική τιμή, δηλαδή το άθροισμα. Η τιμή **all** μπορεί θεωρηθεί ότι αντιπροσωπεύει το σύνολο όλων των τιμών μιας ιδιότητας. Οι εγγραφές με τιμή **all** μόνο για τη διάσταση *color* μπορούν να ανακληθούν από ένα SQL ερώτημα που εκτελεί ένα group by στην στήλη *item-name*. Παρόμοια, μπορεί να χρησιμοποιηθεί ένα group by στο *color* για να πάρει τις εγγραφές με τιμή **all** για το *item-name* και μπορεί να χρησιμοποιηθεί ένα group by χωρίς ιδιότητες (που μπορεί απλώς να παραλειφθεί στην SQL) για να βρει την εγγραφή με τιμή **all** για τα *item-name* και *color*.

<i>item-name</i>	<i>color</i>	<i>number</i>
skirt	dark	8
skirt	pastel	35
skirt	white	10
skirt	all	53
dress	dark	20
dress	pastel	10
dress	white	5
dress	all	35
shirt	dark	14
shirt	pastel	7
shirt	white	28
shirt	all	49
pant	dark	20
pant	pastel	2
pant	white	5
pant	all	27
all	dark	62
all	pastel	54
all	white	48
all	all	164

Εικόνα 3.2 Σχεσιακή αναπαράσταση των δεδομένων της Εικόνας 3.1.

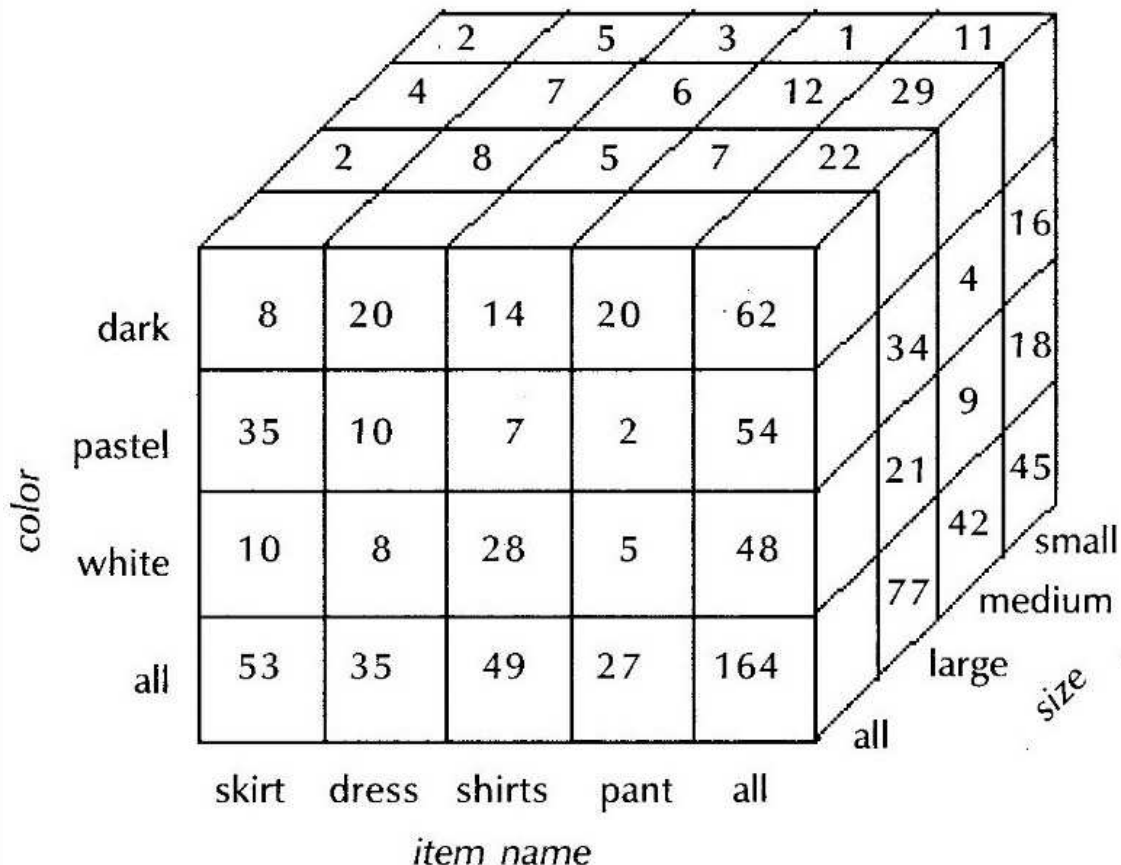
Η γενίκευση ενός δυοδιάστατου πίνακα διασταύρωσης, σε n διαστάσεις, μπορεί να απεικονισθεί οπτικά ως ένας n -διάστατος κύβος, που ονομάζεται **κύβος** δεδομένων. Η Εικόνα 3.3 δείχνει ένα κύβο για τις σχέσεις *sales*. Ο κύβος δεδομένων έχει τρεις διαστάσεις, δηλαδή τις *item-name*, *color* και *size* και η ιδιότητα μέτρησης είναι το *number*. Κάθε κελί προσδιορίζεται με τιμές για αυτές τις τρεις διαστάσεις. Κάθε κελί στον κύβο δεδομένων περιέχει μια τιμή, όπως στον πίνακα διασταύρωσης. Στην Εικόνα 3.3, η τιμή που περιέχεται σε ένα κελί φαίνεται σε μια από τις έδρες του κελιού. Οι άλλες έδρες του κελιού είναι κενές, αν φαίνονται.

Η τιμή για μια διάσταση μπορεί να είναι **all**, οπότε το κελί περιέχει σύνοψη όλων των τιμών αυτής της διάστασης όπως στην περίπτωση των πινάκων διασταύρωσης. Ο αριθμός των διαφορετικών τρόπων με τους οποίους μπορούν να ομαδοποιηθούν οι εγγραφές για σύνοψη μπορεί να είναι μεγάλος. Για ένα πίνακα με n διαστάσεις, η σύνοψη μπορεί να εκτελεστεί με ομαδοποίηση σε καθένα από τα 2^n σύνολα των n διαστάσεων.

Ένα σύστημα online αναλυτικής επεξεργασίας ή OLAP, είναι ένα διαλογικό σύστημα που

επιτρέπει σε ένα αναλυτή να βλέπει μια διαφορετική σύνοψη πολυδιάστατων δεδομένων. Η λέξη *online* δείχνει ότι ένας αναλυτής θα πρέπει να μπορεί να ζητά νέα συνοπτικά δεδομένα και να παίρνει αποκρίσεις online, μέσα σε μερικά δευτερόλεπτα και δεν θα πρέπει να περιμένει πολύ για να δει τα αποτελέσματα ενός ερωτήματος.

Με ένα OLAP σύστημα, ένας αναλυτής μπορεί να κοιτάξει διαφορετικούς πίνακες διασταύρωσης με τα ίδια δεδομένα, επιλέγοντας διαλογικά τις ιδιότητες στο πίνακα διασταύρωσης. Κάθε πίνακας διασταύρωσης είναι μια δυσδιάστατη προβολή ενός πολυδιάστατου κύβου.



Εικόνα 3.3 Τρισδιάστατος κύβος δεδομένων.

Για παράδειγμα, ο αναλυτής μπορεί να επιλέξει ένα πίνακα διασταύρωσης στο *item-name* και *size*, ή ένα πίνακα διασταύρωσης στο *color* και *size*. Η λειτουργία της αλλαγής των διαστάσεων που χρησιμοποιείται σε ένα πίνακα διασταύρωσης ονομάζεται **pivoting** (περιστροφή πίνακα).

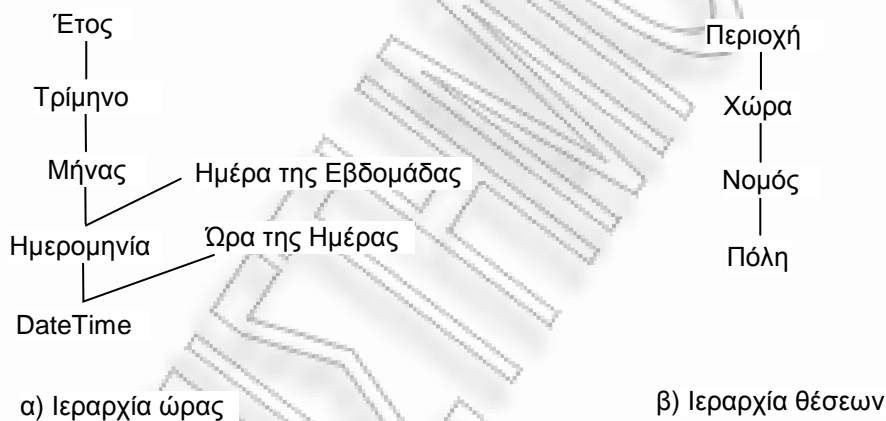
Ένα OLAP σύστημα παρέχει επίσης και άλλη λειτουργικότητα. Για παράδειγμα, ο αναλυτής μπορεί να θέλει να δει ένα πίνακα διασταύρωσης στο *item-name* και *color* για μια σταθερή τιμή του *size*, για παράδειγμα, αντί για το άθροισμα όλων των μεγεθών. Μια τέτοια λειτουργία αναφέρεται ως **τεμαχισμός**, αφού μπορείτε να τη φανταστείτε σαν ένα τεμάχιο από τον κύβο. Η λειτουργία μερικές φορές ονομάζεται **τεμαχισμός σε κύβους**, ιδιαίτερα όταν οι τιμές πολλών διαστάσεων είναι σταθερές.

Όταν χρησιμοποιείται ένας πίνακας διασταύρωσης για προβολή ενός πολυδιάστατου πίνακα, οι τιμές των ιδιοτήτων των διαστάσεων που δεν αποτελούν μέρος του πίνακα διασταύρωσης φαίνονται πάνω από το πίνακα διασταύρωσης. Η τιμή μιας τέτοιας ιδιότητας μπορεί να είναι **all**, όπως φαίνεται στην Εικόνα.3.1, υποδεικνύοντας ότι τα δεδομένα του συνοπτικού πίνακα είναι μια σύνοψη όλων των τιμών για την ιδιότητα. Ο τεμαχισμός / τεμαχισμός σε κύβους

απλώς αποτελείται από την επιλογή συγκεκριμένων τιμών για αυτές τις διαστάσεις, που μετά εμφανίζονται πάνω από τον πίνακα διασταύρωσης.

Τα συστήματα **OLAP** επιτρέπουν στους χρήστες να βλέπουν τα δεδομένα σε **οποιοδήποτε** επίπεδο κατάτμησης. Η λειτουργία της μετακίνησης από τη λεπτότερη κατάτμηση των δεδομένων σε μια πιο άτεχνη κατάτμηση (σε σχέση με το άθροισμα των δεδομένων) ονομάζεται **rollup**. Στο παράδειγμά μας, ξεκινώντας από τον κύβο του πίνακα *sales*, πήραμε το πίνακα **διασταύρωσης** του παραδείγματός μας με **rollup** στη ιδιότητα *size*. Η αντίστροφη λειτουργία δηλαδή, η μετακίνηση από την πιο άτεχνη κατάτμηση στην πιο λεπτή, ονομάζεται **drill down**. Σαφώς, δεν μπορεί να δημιουργηθεί λεπτότερη κατάτμηση από την άτεχνη κατάτμηση. Πρέπει να δημιουργηθεί είτε από τα αρχικά δεδομένα είτε από μια λεπτότερη κατάτμηση συνοπτικών **δεδομένων**.

Οι αναλυτές μπορεί να θέλουν να δουν μια διάσταση σε διαφορετικά επίπεδα **λεπτομέρειας**. Για παράδειγμα, μια ιδιότητα τύπου *datetime* περιέχει την ημερομηνία και την ώρα της ημέρας. Μπορεί να μην έχει έννοια να χρησιμοποιηθεί ακρίβεια χρόνου **δευτερολέπτων ή λιγότερο**. Ένας αναλυτής που ενδιαφέρεται γενικά για το χρόνο, μπορεί να κοιτάξει μόνο στην τιμή της ώρας. Ένας αναλυτής που ενδιαφέρεται για τις πωλήσεις ανά ημέρα της εβδομάδας, μπορεί να απεικονίσει την ημερομηνία σε μια ημέρα της εβδομάδας και να κοιτάξει μόνο αυτό. Ένας άλλος αναλυτής μπορεί να ενδιαφέρεται να συνοψίσει ως προς τον μήνα ή τρίμηνο ή για ένα ολόκληρο έτος.



Εικόνα 3.4 Ιεραρχίες διαστάσεων.

Τα διαφορετικά επίπεδα λεπτομέρειας μιας ιδιότητας μπορούν να οργανωθούν σε μια ιεραρχία. Η Εικόνα 3.4(α) δείχνει μια ιεραρχία στην ιδιότητα *datetime*. Ένα άλλο παράδειγμα, η Εικόνα 3.4(β), δείχνει μια ιεραρχία στη θέση, με την πόλη να είναι στο κάτω μέρος της ιεραρχίας, τον νομό πάνω της, τη χώρα σε επόμενο επίπεδο και τη γενικότερη περιοχή στο κορυφαίο επίπεδο. Στο προηγούμενο παράδειγμα, τα ρούχα μπορούν να ομαδοποιηθούν ως προς την κατηγορία (για παράδειγμα, ρούχα για άντρες ή ρούχα για γυναίκες), το *category* μετά θα βρίσκεται πάνω από το *item-name* στην ιεραρχία για τα ρούχα. Στο επίπεδο των τιμών, οι φούστες και τα φορέματα θα βρίσκονται κάτω από τα ρούχα των γυναικών και τα παντελόνια και τα πουκάμισα κάτω από την κατηγορία με τα ρούχα για άνδρες.

Ένας αναλυτής μπορεί να θέλει να βλέπει τις πωλήσεις των ρούχων χωρισμένες σε ρούχα για γυναίκες και άντρες και να μην ενδιαφέρεται για τιμές. Αφού δει τις συνοπτικές τιμές στο επίπεδο των ανδρών και των γυναικών, μπορεί να θέλει να **κατέβει κάτω στην ιεραρχία για να δει μεμονωμένες τιμές**. Ένας αναλυτής που κοιτάζει στο λεπτομερές επίπεδο μπορεί να **ανέβει επάνω στην ιεραρχία** και να δει ένα πιο γενικό επίπεδο. Και τα δυο επίπεδα μπορούν να εμφανιστούν στον ίδιο πίνακα διασταύρωσης, όπως φαίνεται στην Εικόνα 3.5.

3.2.2 Υλοποίηση OLAP

category	Item-name	dark	pastel	white	total	
womenswear	Skirt	8	8	10	53	88
	dress	20	20	5	35	
	subtotal	28	28	15		
menswear	skirt	14	14	28	49	76
	dress	20	20	5	27	
	subtotal	34	34	33		
total		62	62	48		164

Τα προηγούμενα συστήματα OLAP χρησιμοποιούσαν πολυδιάστατους πίνακες στην μνήμη για αποθηκεύουν κύβους δεδομένων και αναφέρονται ως **πολυδιάστατα OLAP συστήματα** (Multidimensional **OLAP** - **MOLAP**). Αργότερα, τα OLAP ενοποιήθηκαν σε σχεσιακά συστήματα, με δεδομένα αποθηκευμένα σε μια σχεσιακή βάση δεδομένων. Τέτοια συστήματα αναφέρονται ως **σχεσιακά OLAP συστήματα** (relational **OLAP** - **ROLAP**). Υβριδικά συστήματα, που αποθηκεύουν κάποια σύνοψη στην μνήμη και αποθηκεύουν τα βασικά δεδομένα και άλλα συνοπτικά δεδομένα σε μια σχεσιακή βάση δεδομένων, ονομάζονται **υβριδικά συστήματα OLAP (hybrid OLAP - HOLAP)**.

Πολλά OLAP συστήματα υλοποιούνται ως συστήματα πελάτη / διακομιστή. Ο πελάτης περιέχει τη σχεσιακή βάση δεδομένων, όπως επίσης και τους κύβους δεδομένων MOLAP. Τα συστήματα πελάτη παίρνουν προβολές των δεδομένων επικοινωνώντας με τον διακομιστή.

Ένας απλοϊκός τρόπος υπολογισμού ολόκληρου του κύβου (όλων των ομαδοποιήσεων) μιας σχέσης είναι να χρησιμοποιηθεί οποιοσδήποτε λογικός αλγόριθμος για υπολογισμό συνοπτικών λειτουργιών, μια ομαδοποίηση κάθε φορά. Ο απλοϊκός αλγόριθμος θα μπορούσε να πετύχει μεγάλο αριθμό σαρώσεων της σχέσης. Μια απλή οργάνωση είναι να υπολογιστεί μια σχέση, ως πούμε στο (*item-name, color*) από μια σύνοψη του (*item-name, color, size*), αντί από την αρχική σχέση. Για τις τυπικές συνοπτικές συναρτήσεις SQL, μπορούμε να υπολογίσουμε σύνολα με ομαδοποίηση σε ένα σύνολο από ιδιότητες *A* από μια σύνοψη με ομαδοποίηση σε ένα σύνολο από ιδιότητες *B*, αν το $A \subset B$. (Για μερικές μη τυπικές συνοπτικές συναρτήσεις, όπως το median (μέσον), η σύνοψη δεν μπορεί να υπολογιστεί όπως παραπάνω. Η βελτιστοποίηση που περιγράφεται εδώ δεν εφαρμόζεται σε τέτοιες "μη αναλυόμενες" συνοπτικές συναρτήσεις). Η ποσότητα των δεδομένων που διαβάζονται πέφτει κατά πολύ αν υπολογίζουμε μια σύνοψη από μια άλλη σύνοψη, αντί από την αρχική σχέση. Είναι δυνατές και επιπλέον βελτιώσεις. Για παράδειγμα, μπορούν να υπολογιστούν πολλαπλές ομαδοποιήσεις σε μια σάρωση των δεδομένων.

Οι πρώτες υλοποιήσεις OLAP προϋπολόγιζαν και αποθήκευαν ολόκληρους κύβους δεδομένων, δηλαδή ομαδοποιήσεις σε όλα τα υποσύνολα των ιδιοτήτων της διάστασης. Ο προϋπολογισμός επιτρέπει να απαντώνται OLAP ερωτήματα μέσα σε λίγα δευτερόλεπτα, ακόμα και σε σύνολα δεδομένων που μπορεί να περιέχουν εκατομμύρια εγγραφές, με gigabyte από δεδομένα. Ωστόσο, υπάρχουν 2 ομαδοποιήσεις με ιδιότητες *n* διαστάσεων. Οι ιεραρχίες στις ιδιότητες αυξάνουν τον αριθμό περισσότερο. Ως αποτέλεσμα, ολόκληρος ο κύβος δεδομένων είναι συνήθως μεγαλύτερος από την αρχική σχέση που σχημάτισε τον κύβο και σε πολλές περιπτώσεις, δεν είναι εφικτό να αποθηκευτεί ολόκληρος ο κύβος.

Αντί για προϋπολογισμό και αποθήκευση όλων των πιθανών ομαδοποιήσεων, έχει έννοια να γίνει προϋπολογισμός και αποθήκευση μερικών ομαδοποιήσεων και να υπολογίζονται οι άλλοι κατ' απαίτηση. Αντί για υπολογισμό ερωτημάτων από την αρχική σχέση, που μπορεί να χρειαστεί πολύ χρόνο, μπορούμε να τα υπολογίσουμε από άλλα προϋπολογισμένα ερωτήματα. Για παράδειγμα, υποθέστε ότι ένα ερώτημα απαιτεί σύνοψη κατά (*item-name, color*), που δεν έχει προϋπολογιστεί. Το αποτέλεσμα του ερωτήματος μπορεί να υπολογιστεί από σύνοψη των (*item-name, color, Size*), αν αυτό έχει προϋπολογιστεί.

Τα δεδομένα σε ένα κύβο δεδομένων δεν μπορούν να δημιουργηθούν από ένα SQL ερώτημα χρησιμοποιώντας τις βασικές δομές group by, αφού υπολογίζονται συνοπτικές λειτουργίες για διάφορες ομαδοποιήσεις των ιδιοτήτων των διαστάσεων. Η Ενότητα 3.2.3

συζητά τις SQL επεκτάσεις για υποστήριξη λειτουργικότητας OLAP.

3.2.3 Εκτεταμένες Συνοπτικές Λειτουργίες

Η συνοπτική λειτουργικότητα της SQL-92 είναι περιορισμένη, έτσι έχουν υλοποιηθεί διάφορες επεκτάσεις από διαφορετικές βάσεις δεδομένων. Η τυποποίηση SQL:1999 ωστόσο, ορίζει ένα πλούσιο σύνολο από συνοπτικές συναρτήσεις, που περιγράφουμε σε αυτή και στις επόμενες δυο ενότητες. Οι βάσεις δεδομένων Oracle και IBM DB2 υποστηρίζουν τις περισσότερες από αυτές τις λειτουργίες και άλλες βάσεις δεδομένων χωρίς αμφιβολία, θα υποστηρίζουν αυτές τις λειτουργίες στο κοντινό μέλλον.

Οι νέες συνοπτικές λειτουργίες σε ιδιότητες είναι η τυπική απόκλιση και η διακύμανση (std-dev και variance). Η τυπική απόκλιση (standard deviation) είναι η τετραγωνική ρίζα της διακύμανσης (variance). Μερικές βάσεις δεδομένων υποστηρίζουν άλλες συνοπτικές λειτουργίες, όπως το μέσο (median) και την κατάσταση (mode). Μερικές βάσεις δεδομένων επιτρέπουν στους χρήστες να προσθέτουν νέες συνοπτικές, λειτουργίες.

Η SQL :1999 υποστηρίζει επίσης μια νέα κλάση από **δυναμικές συνοπτικές λειτουργίες**, που μπορούν να υπολογίζουν στατιστικά αποτελέσματα σε ζεύγη από ιδιότητες. Περιλαμβάνουν καμπύλες συσχετισμού, του γινομένου της απόκλισης και παλινδρόμησης που δίνουν μια γραμμή προσεγγίζει τη σχέση μεταξύ των τιμών των ζευγαριών των ιδιοτήτων. Οι ορισμοί αυτών των συναρτήσεων μπορούν να βρεθούν σε οποιοδήποτε τυπικό βιβλίο στατιστικής.

Η SQL :1999 υποστηρίζει επίσης γενικεύσεις της δομής group by, χρησιμοποιώντας τις δομές cube και rollup. Μια αντιπροσωπευτική χρήση της δομής cube είναι

```
Select item-name, color, size, sum(number)
from sales
group by cube(item-name, color, size)
```

Αυτό το ερώτημα υπολογίζει την ένωση οκτώ διαφορετικών ομαδοποιήσεων της σχέσης sales:

```
{
(item-name, color, size),(item-name, color),(item-name, size),(color, size),(item-
name),(color),(size),()
}
```

όπου του () δηλώνει μια κενή λίστα group by.

Για κάθε ομαδοποίηση, το αποτέλεσμα περιέχει την κενή τιμή για ιδιότητες που δεν υπάρχουν στην ομαδοποίηση. Για παράδειγμα, ο πίνακας της εικόνας 3.2, με τις επαναλήψεις του all να έχουν αντικατασταθεί με το null, μπορεί να υπολογιστεί από το ερώτημα

```
Select item-name, color, sum(number)
from sales
group by cube (item-name, color)
```

Μια αντιπροσωπευτική δομή **rollup** είναι η

```
select item-name, color, size, sum (number) from sales
group by rollup( item-name, color, size)
```

Εδώ, δημιουργούνται μόνο τέσσερεις ομαδοποιήσεις:

```
{ (item-name, color, size), (item-name, color), (item-name), () }
```

Το `rollup` μπορεί να χρησιμοποιηθεί για να δημιουργήσει συνοπτικά αποτελέσματα σε πολλαπλά επίπεδα μιας ιεραρχίας μιας στήλης. Για παράδειγμα, υποθέστε ότι έχουμε ένα πίνακα `itemcategory(item-name, category)` που δίνει την κατηγορία κάθε στοιχείου. Τότε το ερώτημα

```
select category, item-name, sum (number)
from sales, category
where sales.item-name = itemcategory.item-name
group by rollup (category, item-name)
θα δώσει μια ιεραρχική σύνοψη κατά item-name και category.
```

Πολλαπλά **rollup** και **cube** μπορούν να χρησιμοποιηθούν σε ένα όρο `group by`. Για παράδειγμα, το παρακάτω ερώτημα

```
select item-name, color, size, sum(number) from sales
group by rollup(item-name), rollup(color, size)
```

δημιουργεί τις ομαδοποιήσεις

```
{ (item-name, color, size), (item-name, color), (item-name),
(color, size), (color), () }
```

Ας σημειωθεί ότι το `rollup(item-name)` δημιουργεί δυο ομαδοποιήσεις `{(item-name), ()}` και το `rollup(color, size)` δημιουργεί τρεις ομαδοποιήσεις `{(color, size), (color), ()}`. Το δισταυρούμενο γινόμενο των δυο μας δίνει τις έξι ομαδοποιήσεις που φαίνονται.

Όπως αναφέρθηκε στην Ενότητα 3.2.1, η SQL:1999 χρησιμοποιεί την τιμή **null** για να δείξει την συνηθισμένη σημασία του **null**, όπως επίσης και του **all**. Αυτή η διπλή χρήση του **null** μπορεί να προκαλέσει ασάφεια, αν οι ιδιότητες που χρησιμοποιούνται σε ένα όρο **rollup** ή **cube** περιέχουν κενές τιμές. Η συνάρτηση **grouping** μπορεί να εφαρμοστεί σε μια ιδιότητα και επιστρέφει 1 αν η τιμή είναι μια κενή τιμή που αντιπροσωπεύει το **all** και μηδέν στις άλλες περιπτώσεις. Στο το παρακάτω ερώτημα:

```
select item-name, color, size, sum (number),
grouping(item-name) as item-name-flag,
grouping(color) as color-flag,
grouping(size) as size-flag
from sales
group by cube(item-name, color, size)
```

η έξοδος είναι ίδια με την έκδοση του ερωτήματος χωρίς το **grouping**, αλλά με τρεις επιπλέον στήλες που ονομάζονται `item-name-flag`, `color-flag` και `size-flag`. Σε κάθε εγγραφή, η τιμή ενός πεδίου σημαίας είναι 1, αν το αντίστοιχο πεδίο είναι **null** που αντιπροσωπεύει το **all**.

Αντί να χρησιμοποιήσουμε ετικέτες για να υποδείξουμε ότι τα **null** αντιπροσωπεύουν το **all**, μπορούμε να αντικαταστήσουμε την **null** τιμή με μια τιμή της επιλογής μας:

```
decode (grouping(item-name), 1, 'all', item-name)
```

Αυτή η παράσταση επιστρέφει την τιμή "all" αν η τιμή του *item-name* είναι ένα null που αντιστοιχεί στο **all** και διαφορετικά επιστρέφει την πραγματική τιμή του *item-name*. Αυτή η παράσταση μπορεί να χρησιμοποιηθεί στην θέση του *item-name* στον όρο `select` για να πάρει το "all" στην έξοδο του ερωτήματος, στη θέση των null που αντιπροσωπεύουν το **all**.

Ούτε ο όρος **rollup** ούτε ο **cube** δεν δίνουν πλήρη έλεγχο στις ομαδοποιήσεις που δημιουργούνται. Για παράδειγμα, δεν μπορούμε να τα χρησιμοποιήσουμε για να καθορίσουμε ότι θέλουμε μόνο τις ομαδοποιήσεις {(color, size), (size, item-name)}. Τέτοιες περιορισμένες ομαδοποιήσεις μπορούν να δημιουργηθούν χρησιμοποιώντας την δομή **grouping** στον όρο **having**.

3.2.4 Χαρακτηρισμοί

Η εύρεση της θέσης μιας τιμής σε ένα μεγαλύτερο σύνολο είναι μια συνηθισμένη λειτουργία. Για παράδειγμα, μπορεί να θέλουμε να δώσουμε στους φοιτητές μιας τάξης ένα χαρακτηρισμό ανάλογα με τους συνολικούς βαθμούς τους, με τον χαρακτηρισμό 1 να πηγαίνει στο σπουδαστή με τους υψηλότερους βαθμούς, το 2 στο σπουδαστή με τους επόμενους υψηλότερους βαθμούς κλπ. Ενώ τέτοια ερωτήματα μπορούν να εκφραστούν στην SQL-92, είναι δύσκολο να εκφραστούν και αναποτελεσματικά στον υπολογισμό. Οι προγραμματιστές συνήθως γράφουν το ερώτημα εν μέρει σε SQL και εν μέρει σε μια γλώσσα προγραμματισμού. Ένας σχετικός τύπος ερωτήματος είναι να βρεθεί το ποσοστό στο οποίο ανήκει μια τιμή ενός (πολλαπλού) συνόλου, για παράδειγμα, στο κάτω ένα τρίτο, στο μεσαίο ένα τρίτο ή στο πάνω ένα τρίτο. Μελετούμε εδώ, την υποστήριξη της SQL:1999 για αυτούς τους τύπους των ερωτημάτων.

Ο χαρακτηρισμός γίνεται σε συνδυασμό με μια προδιαγραφή **order by**. Αν έχουμε μια σχέση *student-marks(student-id, marks)* που αποθηκεύει τους βαθμούς κάθε μαθητή, το παρακάτω ερώτημα δίνει τον χαρακτηρισμό σε κάθε σπουδαστή

```
Select student-id, rank () over (order by (marks) desc) as s-rank
From student-marks
```

Παρατηρείται ότι δεν ορίζεται η σειρά των εγγραφών στην έξοδο, έτσι μπορεί να ταξινομηθεί κατά χαρακτηρισμό. Απαιτείται ένας επιπλέον όρος **order by** για να τα βάλει σε ταξινομημένη σειρά, όπως φαίνεται παρακάτω:

```
Select student-id, rank () over (order by {marks} desc) as s-rank from student-marks
order by s-rank
```

Ένα βασικό θέμα με το χαρακτηρισμό είναι πώς να αντιμετωπιστεί στην περίπτωση πολλαπλών εγγραφών που είναι ίδιες στην ιδιότητα διάταξης. Στο παράδειγμά μας, αυτό σημαίνει να αποφασιστεί τι θα συμβεί αν υπάρχουν δυο σπουδαστές με τους ίδιους βαθμούς. Η συνάρτησή **rank** δίνει τον ίδιο χαρακτηρισμό σε όλες τις εγγραφές που είναι ίσες στις ιδιότητες **order by**. Για παράδειγμα, αν ο υψηλότερος χαρακτηρισμός έχει δοθεί σε δυο σπουδαστές και οι δυο θα πάρουν χαρακτηρισμό 1. Ο επόμενος χαρακτηρισμός θα πρέπει να είναι το 3 και όχι το 2, έτσι αν υπάρχουν τρεις σπουδαστές με τον επόμενο υψηλότερο βαθμό, θα πάρουν όλοι χαρακτηρισμό 3 και ο επόμενος φοιτητής θα πάρει 5 κλπ. Υπάρχει επίσης μια συνάρτηση **dense_rank** που δεν δημιουργεί κενά στη διάταξη. Στο παραπάνω παράδειγμα, οι εγγραφές με το δεύτερο υψηλότερο βαθμό θα πάρουν χαρακτηρισμό 2 και οι εγγραφές με τον τρίτο υψηλότερο βαθμό θα πάρουν χαρακτηρισμό 3 κλπ.

Ο χαρακτηρισμός μπορεί να γίνει με τμηματοποίηση των δεδομένων. Για παράδειγμα, έστω ότι έχουμε μια επιπλέον σχέση *student-section(student-id, section)* που αποθηκεύει για κάθε σπουδαστή την ενότητα στην οποία σπουδάζει. Το παρακάτω ερώτημα δίνει το χαρακτηρισμό των σπουδαστών μέσα σε κάθε ενότητα:

```

select student-id, section,
rank () over (partition by section order by marks desc) as sec-rank
from student-marks, student-section
where student-marks.student-id = student-section.student-id order by section, sec-rank

```

Ο εξωτερικός όρος **order by** ταξινομεί τα αποτελέσματα ανά ενότητα και μέσα σε κάθε ενότητα ως προς τον χαρακτηρισμό.

Μπορούν να χρησιμοποιηθούν πολλές παραστάσεις **rank** μέσα σε μια εντολή **select**, έτσι μπορούμε να πάρουμε το συνολικό χαρακτηρισμό και τον χαρακτηρισμό μέσα στην ενότητα χρησιμοποιώντας δυο παραστάσεις **rank** στον ίδιο όρο **select**. Μια ενδιαφέρουσα ερώτηση είναι τι συμβαίνει όταν ο χαρακτηρισμός (πιθανόν με τμηματοποίηση) βρίσκεται μαζί με έναν όρο **group by**. Σε αυτή την περίπτωση, εφαρμόζεται πρώτα ο όρος **group by** και η τμηματοποίηση και ο χαρακτηρισμός γίνονται στα αποτελέσματα του **group by**. Συνεπώς, μπορούν να χρησιμοποιηθούν συνοπτικές τιμές για ένα χαρακτηρισμό. Για παράδειγμα, έστω ότι είχαμε βαθμούς για κάθε φοιτητή, για καθένα από τα διάφορα μαθήματα. Για να χαρακτηρίσουμε τους σπουδαστές από το άθροισμα των βαθμών τους σε διαφορετικά μαθήματα, μπορούμε να χρησιμοποιήσουμε έναν όρο **group by** για να υπολογίσουμε τους συνοπτικούς βαθμούς κάθε σπουδαστή και μετά να χαρακτηρίσουμε τους σπουδαστές με το συνοπτικό άθροισμα.

Μπορούν να χρησιμοποιηθούν συναρτήσεις χαρακτηρισμού για να βρουν τις κορυφαίες n εγγραφές ενσωματώνοντας ένα ερώτημα χαρακτηρισμού μέσα σε ένα ερώτημα εξωτερικού επιπέδου. Παρατηρείται ότι το **bottom n** είναι απλώς το ίδιο με το **top n**, με αντίστροφη σειρά ταξινόμησης. Διάφορα συστήματα βάσεων δεδομένων παρέχουν μη τυπικές επεκτάσεις της SQL για να καθορίζουν κατευθείαν ότι απαιτούνται μόνο τα **top n** αποτελέσματα. Τέτοιες επεκτάσεις δεν απαιτούν την συνάρτηση **rank** και απλοποιούν την δουλειά του βελτιστοποιητή, αλλά αυτή τη στιγμή δεν είναι τόσο γενικές αφού δεν υποστηρίζουν τμηματοποίηση.

Η SQL:1999 καθορίζει επίσης διάφορες άλλες συναρτήσεις που μπορούν να χρησιμοποιηθούν στη θέση του **rank**. Για παράδειγμα, το **percent_rank** μιας εγγραφής δίνει το χαρακτηρισμό της εγγραφής ως κλάσμα. Αν υπάρχουν n εγγραφές στο τμήμα και ο χαρακτηρισμός της εγγραφής είναι r , τότε το ποσοστό ορίζεται ως $(r - 1)/(n - 1)$ (και ως **null** αν υπάρχει μόνο μια εγγραφή στο τμήμα). Η συνάρτηση **cume_dist**, που είναι συντομία για τη συνοπτική κατανομή, μιας εγγραφής ορίζεται ως pin , όπου p είναι ο αριθμός των εγγραφών στο τμήμα, με τιμές διάταξης που είναι μικρότερες ή ίσες με την τιμή διάταξης της εγγραφής και n είναι ο αριθμός των εγγραφών στο τμήμα. Η συνάρτηση **row_number** ταξινομεί τις γραμμές και δίνει σε κάθε γραμμή ένα μοναδικό αριθμό που αντιστοιχεί στη θέση της στη σειρά ταξινόμησης. Διαφορετικές γραμμές με την ίδια τιμή ταξινόμησης θα πάρουν διαφορετικούς αριθμούς γραμμών, με ένα μη προσδιορισμένο τρόπο.

Τέλος, για μια δοθείσα σταθερά n , η συνάρτηση χαρακτηρισμού **ntile(n)** φέρνει τις εγγραφές σε κάθε τμήμα στη συγκεκριμένη σειρά και τις διαιρεί σε n bucket με ίσο αριθμό εγγραφών. Για κάθε εγγραφή, το **ntile(n)** δίνει τον αριθμό του bucket στο οποίο τοποθετείται, με τους αριθμούς των bucket να ξεκινούν από το 1. Αυτή η συνάρτηση είναι πολύ χρήσιμη για κατασκευή ιστογραμμάτων που βασίζονται σε ποσοστά. Για παράδειγμα, μπορούμε να ταξινομήσουμε εργαζόμενους ως προς τον μισθό και να χρησιμοποιήσουμε το **ntile(3)** για να βρούμε σε ποιο εύρος (στο κάτω ένα τρίτο, το μεσαίο ένα τρίτο ή το κορυφαίο ένα τρίτο) ανήκει κάθε εργαζόμενος και να υπολογίσουμε το συνοπτικό μισθό των εργαζομένων σε κάθε εύρος:

```

select threetile, sum (salary)
from (
select salary, ntile(3) over (order by (salary)) as threetile from employee) as s
group by threetile.

```

Η παρουσία των κενών τιμών μπορεί να περιπλέξει τον ορισμό του χαρακτηρισμού, αφού δεν είναι σαφές πού θα βρίσκονται στη σειρά ταξινόμησης. Η SQL: 1999 επιτρέπει στο χρήστη να καθορίζει πού θα πρέπει να βρίσκονται, χρησιμοποιώντας το **nulls first** ή **nulls last**. Για παράδειγμα:

```

Select student-id, rank () over (order by marks desc nulls last) as s-rank
From student-marks

```

3.2.5 Παράθυρα

Ένα παράδειγμα ενός ερωτήματος παραθύρου είναι το ερώτημα που αν δοθούν οι τιμές πωλήσεων για κάθε ημερομηνία, υπολογίζει για κάθε ημερομηνία το μέσο όρο των πωλήσεων αυτής της ημέρας, της προηγούμενης ημέρας και της επόμενης ημέρας. Τέτοια ερωτήματα μετακινούμενου μέσου όρου χρησιμοποιούνται για να ομαλοποιούν τις τυχαίες αποκλίσεις. Ένα άλλο παράδειγμα ενός ερωτήματος παραθύρου είναι αυτό που βρίσκει το συνοπτικό υπόλοιπο ενός λογαριασμού, αν δοθεί μια σχέση που καθορίζει τις καταθέσεις και τις αναλήψεις ενός λογαριασμού. Τέτοια ερωτήματα είτε είναι δύσκολο είτε είναι αδύνατον (ανάλογα με το ερώτημα) να εκφραστούν στην βασική SQL.

Η SQL:1999 παρέχει μια λειτουργία παραθύρου για υποστήριξη τέτοιων ερωτημάτων. Αντίθετα με το **group by**, η ίδια εγγραφή μπορεί να υπάρχει σε πολλά παράθυρα. Αν υποθέσουμε ότι παίρνουμε μια σχέση *transaction(account-number, date-time, value)*, όπου το *value* είναι θετικό για μια κατάθεση και αρνητικό για μια ανάληψη και ότι υπάρχει το πολύ μια συναλλαγή ανά τιμή *date-time*, προκύπτει το ερώτημα:

```

Select account-number, date-time, sum {value} over
(partition by account-number order by date-time rows unbounded preceding) as
balance from transaction
Order by account-number, date-time

```

Το ερώτημα δίνει τα αθροιστικά υπόλοιπα κάθε λογαριασμού ακριβώς πριν από κάθε συναλλαγή του λογαριασμού. Το αθροιστικό υπόλοιπο του λογαριασμού είναι το άθροισμα των τιμών όλων των προηγούμενων συναλλαγών του λογαριασμού.

Ο όρος **partition by** τμηματοποιεί τις εγγραφές κατά αριθμό λογαριασμού, έτσι για κάθε γραμμή εξετάζονται μόνο οι εγγραφές στο τμήμα της. Δημιουργείται ένα παράθυρο για κάθε εγγραφή. Οι λέξεις κλειδιά **rows unbounded preceding** καθορίζουν ότι το παράθυρο κάθε εγγραφής αποτελείται από όλες τις εγγραφές του τμήματος που είναι πριν από αυτές στη συγκεκριμένη σειρά (εδώ, αυξάνοντας τη σειρά του *date-time*). Η συνοπτική συνάρτηση **sum(value)** εφαρμόζεται σε όλες τις εγγραφές του παραθύρου. Το ερώτημα δεν χρησιμοποιεί ένα όρο **group by**, αφού υπάρχει μια εγγραφή εξόδου για κάθε εγγραφή στη σχέση *transaction*.

Ενώ το ερώτημα θα μπορούσε να γραφτεί χωρίς αυτές τις δομές, θα ήταν μάλλον δύσκολο να σχηματιστεί. Μπορεί επίσης να επικαλύπτονται διάφορα παράθυρα, δηλαδή μια εγγραφή μπορεί να υπάρχει σε περισσότερα από ένα παράθυρα.

Μπορούν να καθοριστούν και άλλοι τύποι παραθύρων. Για παράδειγμα, για να πάρουμε ένα παράθυρο που περιέχει τις προηγούμενες δέκα γραμμές κάθε εγγραφής, μπορούμε να καθορίσουμε το **rows 10 preceding**. Για να πάρουμε ένα παράθυρο που περιέχει την τρέχουσα, την προηγούμενη και την επόμενη εγγραφή, μπορούμε να χρησιμοποιήσουμε το **between rows 1 preceding and 1 following**. Για να πάρουμε τις προηγούμενες γραμμές και την τρέχουσα γραμμή, μπορούμε να πούμε **between rows unbounded preceding and current**. Αν η διάταξη είναι μια ιδιότητα χωρίς κλειδί, το αποτέλεσμα δεν είναι προσδιοριστικό, αφού η σειρά των εγγραφών δεν ορίζεται πλήρως.

Μπορούμε ακόμα να καθορίσουμε παράθυρα κατά εύρος τιμών, αντί κατά αριθμούς γραμμών. Για παράδειγμα, έστω ότι η τιμή διάταξης μιας εγγραφής είναι *v*. Τότε, το **range between 10 preceding and current row** θα δώσει τιμές των οποίων η τιμή διάταξης είναι μεταξύ *v-10* και *v* (συμπεριλαμβανομένων και αυτών των τιμών). Όταν χειριζόμαστε ημερομηνίες, μπορούμε να χρησιμοποιήσουμε το **range interval 10 day preceding** για να πάρουμε ένα παράθυρο που περιέχει εγγραφές μέσα στις προηγούμενες δέκα ημέρες, αλλά όχι συμπεριλαμβανομένης της ημερομηνίας της εγγραφής.

Διαπιστώνεται λοιπόν η λειτουργικότητα των παραθύρων της SQL:1999, η οποία είναι

πολύ πλούσια και μπορεί να χρησιμοποιηθεί για να διατυπωθούν μάλλον περίπλοκα ερωτήματα, με μικρή προσπάθεια.

4. Γεωγραφικά Συστήματα Πληροφοριών

4.1 Ιστορικό

Οι προτάσεις της τεχνολογίας των πληροφοριακών συστημάτων για την αντιμετώπιση ζητημάτων χωρικής ανάλυσης και σχεδιασμού έχουν τις ρίζες τους στα μέσα της δεκαετίας του '60, όταν εφαρμόστηκε για πρώτη φορά σε εντατικό και πρακτικό επίπεδο η ιδέα της οργάνωσης και συστηματοποίησης της γεωγραφικής πληροφορίας μέσω ηλεκτρονικού υπολογιστή. Η αρχική σύλληψη της συνδυασμένης χρήσης θεματικών κυρίως χαρτών για τον προσδιορισμό θέσης και χωρικών συσχετισμών ανάμεσα σε στοιχεία και φαινόμενα του γεωγραφικού χώρου, είναι αρκετά παλαιότερη της ανακάλυψης των ηλεκτρονικών υπολογιστικών συστημάτων και η ιστορία της χαρτογραφίας έχει να επιδείξει πολλά παραδείγματα ηλικίας αρκετών αιώνων.

Οι πρώτες σύγχρονες συστηματικές απόπειρες συνδέονται με τις τεχνικές και μεθοδολογίες της χαρτογραφίας και καταγράφονται στα πεδία της πολεοδομίας και της χωροταξίας στη Γερμανία, τις Ηνωμένες Πολιτείες της Αμερικής και τη Μεγάλη Βρετανία. Χαρακτηριστικά παραδείγματα αποτελούν οι πολεοδομικές καταγραφές και η χαρτογράφηση της εξέλιξης της περιφέρειας του Dusseldorf [1912], κυκλοφοριακές καταγραφές και σχέδια χρήσεων γης στην πόλη Massachusetts κατά την ίδια χρονολογία, χαρτογραφικές σειρές τοπογραφικών αποτυπώσεων σε συνδυασμό με τις γενικές χρήσεις γης και των δυνατοτήτων πρόσβασης στον αστικό χώρο της περιοχής Doncaster της Αγγλίας κατά το 1922. Χαρακτηριστική είναι η χαρτογραφική σειρά «Survey of New York and Its Environs» που εκδόθηκε το 1929. Στη σειρά αυτή γίνεται για πρώτη φορά αξιοποίηση διαδικασιών αλγεβρικών πράξεων μεταξύ χαρτών. Οι πράξεις αυτές βασίζονται στην κατακόρυφη υπέρθεση μέσω χαρτογραφικών συντεταγμένων και αποτελούν έως σήμερα την ισχυρότερη ίσως γεωμετρική δυνατότητα των σύγχρονων λογισμικών ΣΓΠ.

Στη βρετανική έκδοση του 1950 «Town and Country Planning Textbook» περιέχεται οργανωμένη σύνθεση πολλαπλών χαρτογραφικών επιπέδων που αφορούν σε εκτεταμένες χωρικές και χρονικές αναφορές και σε τέσσερις κατηγορίες δεδομένων: υψομετρία εδάφους, γεωλογία, υδρολογία / ύδρευση, και αγροτικές χρήσεις γης. Στην έκδοση περιγράφεται αναλυτικά η μεθοδολογία των αναγωγών σε ενιαία χαρτογραφική κλίμακα, της χαρτογραφικής γενίκευσης, των χρονικών αναγωγών, αλλά και των διαδικασιών χαρτογραφικών πράξεων μεταξύ των πληροφοριακών επιπέδων που χρησιμοποιούνται. Ακολούθως, κατά το έτος 1962, ομάδα πολεοδόμων στο Massachusetts Institute of Technology συνέβαλε σημαντικά στην εξέλιξη των χαρτογραφικών πράξεων με τη χρήση 26 σταθμισμένων χαρτογραφικών επιπέδων πληροφορίας για τον προσδιορισμό της κυκλοφοριακής δομής της ευρύτερης περιοχής της πόλης. Η συνδυαστική χαρτογραφική τεχνική που χρησιμοποίησαν βασιζόταν κυρίως σε φωτογραφικές μεθόδους αναπαράστασης των συνθετικών πληροφοριακών επιπέδων. Κατά τη διάρκεια της ίδιας δεκαετίας, άρχισαν να εκδίδονται συστηματικά οι πρώτοι θεματικοί χάρτες συνδυασμένων πληροφοριακών επιπέδων σε τυποποιημένες κλίμακες και με μηχανικές μεθόδους, οι οποίοι αφορούσαν κυρίως σε στοιχεία τοπογραφίας, γεωλογίας και κάλυψης γης από υπηρεσίες των ΗΠΑ (U.S. Geological Survey και Natural Resource Conservation Service).

Το σκηνικό στο χώρο της χαρτογραφίας άλλαξε θεαματικά με την εμφάνιση των πρώτων υπολογιστικών συστημάτων. Το 1959, στο περιοδικό «Geographical Review» γίνεται η πρώτη δημοσίευση σχετικά με την ανάπτυξη ενός απλού μοντέλου χαρτογραφικών εφαρμογών με χρήση ενός πρωτόγονου ηλεκτρονικού υπολογιστή. Η δημοσίευση παρουσίαζε ένα χαρτογραφικό σύστημα βασισμένο σε τρία βασικά συστατικά: «map input - map manipulation - map output» (είσοδος - διαχείριση - έξοδος χάρτη). Τα τρία αυτά απλά βήματα αποτελούν την καταβολή της «γεω-κωδικοποίησης - καταχώρισης - αποθήκευσης δεδομένων», της «διαχείρισης - ανάλυσης δεδομένων» και της «χαρτογραφικής αναπαράστασης» που αποτελούν έως και σήμερα τον πυρήνα των σύγχρονων λογισμικών ΣΓΠ. Στη συνέχεια εμφανίζεται σειρά λογισμικών προγραμμάτων βασισμένων στη γλώσσα προγραμματισμού «fortran», με τα οποία γίνονται οι πρώτες προσπάθειες ψηφιοποίησης χαρτογραφικών

δεδομένων με τη χρήση περιφερειακών συσκευών (digitizers) καθώς επίσης και ψηφιακής χαρτογραφικής απόδοσης με τη χρήση εκτυπωτών και σχεδιογράφων.

Βαθμηδόν, τα απλά προγράμματα άρχισαν να αντικαθίστανται από οργανωμένες συλλογές λογισμικού (software packages) με προσανατολισμό το ψηφιακό σχέδιο και τη χαρτογραφική αναπαράσταση που γίνονται γνωστά με τα ονομαστικά: «computer aided design - CAD», «computer aided mapping - CAM» και «computer aided cartography - CAC». Τα συστήματα αυτά επιλύουν προβλήματα ταχύτητας, ακρίβειας, μετατροπών ανάμεσα σε χαρτογραφικές κλίμακες, απλών χαρτογραφικών πράξεων και χαρτογραφικής αναπαραγωγής. Ορισμένα συστήματα του είδους λειτούργησαν για αρκετά χρόνια στις ΗΠΑ, εξοπλίζοντας πολλές δημόσιες υπηρεσίες και προσανατολιζόνταν κυρίως στην ανάλυση, διαχείριση και σύνθεση χαρτογραφικών πληροφοριακών επιπέδων και την παραγωγή χωροπληθών και ισარიθμικών χαρτών. Μία δεύτερη σειρά λογισμικών συλλογών από την οποία θεωρείται πως τα ΣΓΠ έλκουν επίσης μέρος της καταγωγής τους, είναι εκείνη των συστημάτων διαχείρισης βάσεων δεδομένων (data base management systems - DBMS) και ιδίως των συστηματικών χαρτογραφικών βάσεων δεδομένων. Η πρώτη, που αναπτύχθηκε από γνωστή υπηρεσία πληροφοριών των ΗΠΑ, ήταν μία παγκόσμια «τράπεζα δεδομένων» η οποία αφορούσε σε στοιχεία ακτογραμμών, υδρολογίας και εθνικών και άλλων διοικητικών συνόρων, με γεωγραφική κάλυψη ολόκληρου του πλανήτη. Εκδόθηκε σε διάφορες χαρτογραφικές προβολές και κλίμακες, στηρίχθηκε σε λογισμικές συλλογές προγραμμάτων «CAM», ορισμένα δε στοιχεία της χρησιμοποιούνται ακόμη και σήμερα.

Κατόπιν σειράς προσπαθειών σύνθεσης των δύο ανωτέρω εξελίξεων (των συστημάτων διαχείρισης της γεωμετρικής πληροφορίας και των περιγραφικών βάσεων δεδομένων), οι οποίες απέδωσαν πολλά πρωτότυπα συστήματα, παρουσιάστηκε το σύστημα «DIME» (dual independent map encoding) ως πειραματική επινόηση της στατιστικής υπηρεσίας των ΗΠΑ (U.S. Census Bureau). Το σύστημα έδινε ως πληροφοριακό αποτέλεσμα τα επονομαζόμενα ως «αρχεία γεωγραφικών βάσεων δεδομένων» (geographic base files - GBF) τα οποία είχαν ως βασική αρχή τη σύνδεση ανάμεσα στην γραφική-γεωμετρική με την περιγραφική-στατιστική πληροφορία και θεωρήθηκαν ως ιστορική τομή στο χώρο της διαχείρισης και αναπαράστασης της γεωγραφικής πληροφορίας. Με τον τρόπο αυτό ήταν δυνατόν να αποδίδεται χαρτογραφικά όλο το συλλεγμένο απογραφικό υλικό, να υποβοηθείται ο σχεδιασμός των απογραφικών διαδικασιών, αλλά και να επιτρέπονται αναζητήσεις (spatial queries) σε επίπεδο χωρικών προτύπων και κατανομών.

Ακριβώς κατά την περίοδο αυτή, στα μέσα της δεκαετίας του '60, αναπτύχθηκε σε ολοκληρωμένο επίπεδο η κατά πολλούς δεύτερη σπουδαιότερη, μετά το χάρτη, τεχνολογική ανακάλυψη της Γεωγραφίας: τα ΣΓΠ (Geographical Information Systems - G.I.S.). Η τεχνολογία αυτή γνωρίζει έκτοτε μία ευρεία σειρά εφαρμογών, σχεδόν σε κάθε ζήτημα ανάλυσης και σχεδιασμού όπου η παράμετρος «Γεωγραφικός Χώρος» υπεισέρχεται άμεσα ή έμμεσα (ζητήματα χωροταξίας, αστικής και περιφερειακής ανάπτυξης, διαχείρισης φυσικών πόρων, περιβαλλοντικών μελετών, πολεοδομικού σχεδιασμού, κτηματολογίου κ.λπ.). Ορισμένα από τα πρώτα συστήματα τα οποία θεωρούνται ως ορόσημα του χώρου αναπτύχθηκαν στον Καναδά (Canada Geographic Information System, 1964) και τις Η.Π.Α. (Minnesota Land Management Information System, 1969 - εστίαζε σε πληροφορίες λεπτομερούς επιπέδου και κτηματολογικού κυρίως περιεχομένου- και «Land Use and Natural Resources», Νέα Υόρκη, 1967 - προσανατολισμένο σε περιβαλλοντικά κυρίως ζητήματα).

Προς το τέλος της δεκαετίας, στο πανεπιστήμιο του Harvard λειτούργησε το εργαστήριο «Laboratory for Computer Graphics and Spatial Analysis» στο οποίο επιτελέστηκαν τα πρώτα σημαντικά θεωρητικά βήματα, και αναπτύχθηκαν αρκετά ολοκληρωμένα συστήματα, με σημαντικότερο από αυτά το σύστημα «Odyssey». Στο σύστημα αυτό αναπτύσσεται ένα πρωτοποριακό μοντέλο τοπολογικής ολοκλήρωσης των αποθηκευόμενων χαρτογραφικών στοιχείων, το οποίο έτυχε ευρύτατης χρήσης στα περισσότερα λογισμικά ΣΓΠ μετά τη δημοσίευσή του το 1975. Το μοντέλο ονομάστηκε «arc/node» ή «vector data structure» και βασίστηκε σε αλγοριθμικές ρουτίνες που ταξινομούν τις σειρές ψηφιοποίησης ως «γραμμές» και «αλυσίδες γραμμών», τις οποίες συναθροίζουν τοπολογικά δημιουργώντας τις πολυγωνικές οντότητες οποίες «εκπροσωπούν» μέσα στο σύστημα τα επιφανειακά χαρτογραφικά δεδομένα. Οι γραμμές (arcs) δομούνται ως στοιχειώδη ευθύγραμμα διανυσματικά τμήματα (vectors) με έναν «κόμβο» (node) αρχής και έναν τέλος με γνωστές οριζοντιογραφικές συντεταγμένες. Τα «πολύγωνα» (polygons) συντίθενται ως κλειστή διαδοχή τόξων και οι ρουτίνες τοπολογικής

δόμησης που εκτελούνται εν συνεχεία εξασφαλίζουν μέσω ειδικών πινάκων την πληροφορία γεϊντίασης και σύνδεσης ανάμεσα στα γεωμετρικά χαρακτηριστικά.

Το 1969 ιδρύθηκε η ESRI στην Καλιφόρνια των ΗΠΑ από ένα σπουδαστή του εργαστηρίου του Harvard, τον Jack Dangermond. Η εταιρία ανέπτυξε τα πρώτα λογισμικά προγράμματα και οργανωμένες συλλογές προγραμμάτων με την ονομασία «Arc/Info», ταυτόχρονα με την ανάληψη των πρώτων μεγάλων προγραμμάτων και εφαρμογών ΣΓΠ, δημιουργώντας το πλέον αναγνωρίσιμο, έως και σήμερα εμπορικό «πακέτο» διαχείρισης γεωγραφικής πληροφορίας.

Το 1974 η Διεθνής Γεωγραφική Ένωση (International Geographic Union) εξέδωσε την πρώτη καταγραφή οργανωμένων λογισμικών προγραμμάτων G.I.S. σε έναν τόμο ο οποίος τιτλοφορήθηκε «Complete Geographical Information Systems». Καθώς τα πρώτα χρόνια δεν υπήρχε κάποια οριοθέτηση ή ενιαία ορολογία για την περιγραφή των συστημάτων γεωγραφικών πληροφοριών και για τις λειτουργίες τους, η αναφορά αυτή αποτέλεσε την αφετηρία σύγκλισης στο χώρο των G.I.S. τόσο σε ερευνητικό επίπεδο, όσο και σε επίπεδο εφαρμογών. Την εποχή εκείνη τέθηκε για πρώτη φορά ο προβληματισμός σε σχέση με την επικάλυψη ανάμεσα στα χαρτογραφικά συστήματα και τα συστήματα γεωγραφικών πληροφοριών. Ο προβληματισμός αυτός εξακολουθεί να υπάρχει έως σήμερα, καθόσον σε αρκετούς επιστημονικούς και μη κύκλους εξακολουθεί να υφίσταται σύγχυση στην οριοθέτηση των δύο πεδίων. Ο Kurt Brassel σημειώνει: «Αντιλαμβανόμεθα ότι ένα χαρτογραφικό σύστημα σχεδιάζεται κυρίως για σκοπούς χαρτογραφικής αναπαράστασης της πραγματικότητας, ακόμη και εάν συμπεριλαμβάνει δευτερεύουσες "μη-γραφικές" (non-graphical) λειτουργίες. Ένα σύστημα γεωγραφικής πληροφορίας σχεδιάζεται για ένα ευρύτερο φάσμα εφαρμογών, ακόμη και εάν οι χαρτογραφικές του λειτουργίες αποτελούν ένα σημαντικό υποσύνολο των δυνατοτήτων του. Τα ΣΓΠ και τα ψηφιακά χαρτογραφικά συστήματα έχουν και θα εξακολουθήσουν να έχουν αυτή τη σημαντική δομική επικάλυψη στο περιεχόμενο τους».

Η ανάπτυξη των ΣΓΠ παρέμεινε σταθερή κατά τη δεκαετία του 1980 χρησιμοποιώντας μεγάλα υπολογιστικά συστήματα (mainframes). Ταυτόχρονα, έκαναν την εμφάνισή τους οι πρώτοι μικρό-υπολογιστές (microcomputers) και τούτο αποτέλεσε μία επαναστατική εξέλιξη στη διάδοση των συστημάτων, αφού άμεσα οι μεγαλύτερες εταιρίες λογισμικού ΣΓΠ άρχισαν τη δύσκολη μετάβαση προς τους προσωπικούς ηλεκτρονικούς υπολογιστές περνώντας αρχικά από το στάδιο των «σταθμών εργασίας» (workstations) και τα λειτουργικά συστήματα τύπου «UNIX».

Η εποχή αυτή, έως και τις αρχές της δεκαετίας του 1990, θεωρείται ως εποχή «ωρίμανσης» της τεχνολογίας των ΣΓΠ. Ο εξοπλισμός κοστίζει λιγότερο, τα προγράμματα γίνονται εύχρηστα, η σύνδεση των επιστημονικών κοινοτήτων γίνεται απλούστερη μέσω του διαδικτύου, ενώ δημιουργείται μία πλειάδα επιστημονικών ενώσεων, αντίστοιχων πανεπιστημιακών τμημάτων, δημόσιων και ιδιωτικών οργανισμών και φυσικά μεγάλου αριθμού εφαρμογών. Η ίδια εποχή, πλησιάζοντας προς τον παρόντα χρόνο, χαρακτηρίζεται από πολλούς ως η εποχή της «Γεωπληροφορικής», αφού υπάρχουν εκτιμήσεις ότι το 80% όλων των ψηφιακών στοιχείων τα οποία έχουν παραχθεί παγκοσμίως, περιλαμβάνουν κάποιο είδος γεωγραφικής αναφορά (π.χ. γεωγραφικές συντεταγμένες, διευθύνσεις, ταχυδρομικοί κώδικες, κ.λπ.). Αυτή η αναφορά στο γεωγραφικό χώρο επιβάλλει και επιτρέπει την ολοκλήρωση μίας μεγάλης σειράς ετερόκλητων πληροφοριών. Ταυτόχρονα, το μέγεθος και η πολυπλοκότητα των στοιχείων τα οποία μπορούν να οργανωθούν μέσα σε ένα κοινό πλαίσιο, χρησιμοποιώντας ακριβώς τις κοινές γεωγραφικές τους συνδέσεις, παρουσιάζουν μία εξαιρετική πρόκληση για τις επιστήμες της πληροφορίας: «πώς μετασχηματίζουμε αυτά τα στοιχεία σε πληροφορίες, και στη συνέχεια τις πληροφορίες σε γνώση» πλησιάζοντας (έστω με αρκετή δόση υπερβολής) στην «Κοινωνία των Γεωγραφικών Πληροφοριών» (Geographical Information Society).

Στην Ελλάδα η δεκαετία 1988 - 1998, χαρακτηρίζεται από την είσοδο της τεχνολογίας των Συστημάτων Γεωγραφικών Πληροφοριών στα τεχνικά δρώμενα της χώρας. Η δεκαετία αυτή μπορεί να χαρακτηρισθεί ως μία περίοδος εντός της οποίας μία σχετικά νέα τεχνολογία παρείσφρησε στην καθημερινότητα των εμπλεκομένων με τη γεωγραφική πληροφορία, κάνοντας κάποιους από αυτούς εκστατικούς εξ' αιτίας της όποιας πολυπλοκότητας των συστημάτων αυτών, κάποιους ευτυχισμένους που ανακάλυψαν επί τέλους αντικειμενικά εργαλεία που τους αποδέσμευαν από έναν υποκειμενισμό στη διαχείρισή της και ταυτόχρονα τους προσέφεραν πολλές εναλλακτικές δυνατότητες στη λήψη αποφάσεων και κάποιους άλλους, είτε απλώς επιφυλακτικούς είτε να χάνουν στην κυριολεξία τον ύπνο τους, καθώς

βρέθηκαν σε μία κατάσταση ανεξέλεγκτη γι' αυτούς, αφού το βασικό συστατικό των συστημάτων αυτών ήταν η πληροφορική της οποίας η καθημερινή χρήση αντιμετωπιζόταν ως καινοφανής δαίμονας.

Ουσιαστικά, τίποτε από τα παραπάνω δεν κυριάρχησε. Το σημαντικό γεγονός ωστόσο, ήταν ότι η αναταραχή αυτή δημιούργησε μία νέα δυναμική στις συνιστώσες της χρήσης και της διαχείρισης της γεωγραφικής πληροφορίας. Η σε πρωταρχικό στάδιο ερευνητικής φύσης αντιμετώπιση της «μοντέρνας αυτής στροφής» της κοινότητας, έδωσε σταδιακά τη θέση της σε διαδικασίες και πρακτικές αντιμετώπισης και ενσωμάτωσης τελικά της τεχνολογικής αυτής στροφής στην καθημερινότητα. Νέα εκπαιδευτικά αντικείμενα θεμελιωμένα στον τρόπο τις αρχές και τις μεθοδολογίες διαχείρισης της γεωγραφικής πληροφορίας έκαναν -στην αρχή δειλά αλλά στη συνέχεια πολύ δυναμικά- την εμφάνισή τους στα προγράμματα σπουδών της Τριτοβάθμιας Εκπαίδευσης, νέα Τμήματα ιδρύθηκαν εξ αιτίας ακριβώς αυτής της δυναμικής, ο ιδιωτικός τομέας αργά αλλά σταθερά στελεχώθηκε από ανθρώπους ικανούς να διαχειρίζονται την νέα αυτή τεχνολογία παράγοντας νέα προϊόντα και τέλος, το δημόσιο, είτε ως κεντρική είτε ως τοπική διοίκηση ενσωμάτωσε στα οργανογράμματα των υπηρεσιών του την τεχνολογία αυτή, καθώς γρήγορα θα καλείτο να θεμελιώσει μεγάλο μέρος της παραγωγικής του διαδικασίας δια μέσου αυτής. Πράγματι το 1994 το τότε Υπουργείο Γεωργίας προκήρυξε το Ολοκληρωμένο Σύστημα Διαχείρισης και Ελέγχου (ΟΣΔΕ) και ο ΟΚΧΕ το 1996 το Εθνικό Κτηματολόγιο, έργα τα οποία όχι μόνο απορρόφησαν εξειδικευμένο σε ΣΓΠ επιστημονικό δυναμικό, αλλά υπέδειξαν επίσης έντονα και την έλλειψή του. Ειδική μνεία πρέπει να γίνει στην Ελληνική Εταιρία Γεωγραφικών Συστημάτων Πληροφοριών (HellasGIS) που ιδρύθηκε το 1998 με στόχο την προώθηση της τεχνολογίας, της επιστήμης και των εφαρμογών ΣΓΠ, μέσω διοργάνωσης διαλέξεων, σεμιναρίων, εκθέσεων, πανελληνίων και διεθνών συνεδρίων, ανάπτυξης της έρευνας και την συμμετοχή σε ερευνητικά προγράμματα, την έκδοση σχετικού περιοδικού (Γεωανάλεκτα) κ.ά.

Στην επόμενη δεκαετία (1999 - 2009), όπως ήταν φυσικό πολλά πράγματα άλλαξαν. Η μεγέθυνση της κοινότητας με όλο και περισσότερους ενασχολούμενους με τη γεωγραφική πληροφορία μέσω της τεχνολογίας των ΓΣΠ, είχε ως αντανάκλαστικό μέγεθος αλλαγές στην παραγωγική διαδικασία. Ο όρος «Γεωπληροφορική» προστέθηκε στο «τρέχον» λεξιλόγιο της κοινότητας των ενασχολούμενων με τη γεωγραφική πληροφορία, καθώς έπρεπε να δηλώνεται προφανώς η τεχνολογική συσχέτιση Γεωγραφίας (με την ευρεία έννοια) και Πληροφορικής, και το τοπίο στην παραγωγική διαδικασία μελετών που αφορούν με κάθε τρόπο το γεωγραφικό χώρο άλλαξε δραματικά. Οι συνέπειες της δυναμικής αυτής φάνηκαν μέσα από τις προηγηθείσες και προαναφερθείσες προκηρύξεις μεγάλων έργων Γεωπληροφορικής από φορείς του δημοσίου αλλά και ότι από 'κει και πέρα ακολούθησε.

Όπως ήταν αναμενόμενο, η Ελληνική κοινότητα των Γεωγραφικών Πληροφοριών βρέθηκε σε ένα νέο τοπίο σε ότι αφορούσε πλέον την «τάξη» της παραχθείσας και παραγόμενης γεωγραφικής πληροφορίας. Μ' άλλο λόγια, τις μεθόδους και τα πρότυπα της παραγωγής της, τα πρότυπα των αποτελεσμάτων από την παραγωγή της, πρακτικές, μεθόδους και πρότυπα της διαχείρισης, αποθήκευσης και διάθεσής της και όλα αυτά μέσα σε ένα ταχύτατα εξελισσόμενο πληροφορικό περιβάλλον όπου η διάχυση των διαδικτυακών διαδικασιών είχε ήδη προχωρήσει και είχε δημιουργήσει διαφορετικές λογικές και απαιτήσεις.

Παράλληλα, ο σχετικά νέος φορέας του δημοσίου της Κοινωνίας της Πληροφορίας εκμεταλλευόμενος τα κοινοτικά κονδύλια του Γ' ΚΠΣ σχεδίαζε πολλά και μεγάλα έργα Γεωπληροφορικής τα οποία επρόκειτο σταδιακά να βγουν στο κοινό για διαβούλευση. Εκ παραλλήλου με τη συνέχιση των προαναφερθέντων έργων του ΟΣΔΕ, το τότε Υπουργείο Γεωργίας συνεχίζοντας την καλή παράδοση που πρώτο αυτό ξεκίνησε, προκήρυξε τα έργα του Ελαιοκομικού κα; Αμπελοουργικού Μητρώου και τις εφαρμογές της Τηλεπισκόπησης στη Γεωργία για τη χώρα και αρκετοί ΟΤΑ έβαλαν στις καθημερινές δραστηριότητες τους διαδικασίες γεωπληροφορικής. Ταυτόχρονα, μία οδηγία της Ευρωπαϊκής Επιτροπής που αφορούσε στη διαχείριση των υδατικών πόρων των κρατών μελών της Ευρωπαϊκής Ένωσης, η 2000/60, επέρχεται για να υποδείξει τον κυρίαρχο ρόλο της διαχείρισης της γεωγραφικής πληροφορίας και σε αυτόν τον πράγματι ευαίσθητο τομέα.

Την τριετία 2002-2004 η Κοινωνία της Πληροφορίας μέσα από τη δράση «Περιφερειακά Γεωγραφικά Συστήματα και Καινοτόμες Ενέργειες» έχει ήδη προβεί σε προκηρύξεις έργων που αφορούν τόσο την κεντρική όσο και την περιφερειακή διοίκηση αλλά και την τοπική αυτοδιοίκηση, το Υπουργείο Ανάπτυξης ετοιμάζει τη διαβούλευση της προκήρυξης για την

εφαρμογή των ΓΣΠ στη διαχείριση υδατικών πόρων, μεγάλοι δημόσιοι οργανισμοί χαρτογραφικής παραγωγής όπως ο ΟΚΧΕ, η ΓΥΣ, το ΙΓΜΕ και το ΥΠΓε ετοιμάζονται μέσω χρηματοδότησης από την παραπάνω δράση της Κοινωνίας της Πληροφορίας, να συμβάλλουν με προκηρύξεις συγκεκριμένων έργων στην ίδρυση μίας εθνικής υποδομής γεωγραφικών δεδομένων, προετοιμάζονται ειδικά συστήματα βασισμένα στην τεχνολογία των ΓΣΠ για την ασφάλεια των Ολυμπιακών Αγώνων της Αθήνας και γενικότερα η διάχυση και η διεύθυνση της Γεωπληροφορικής στην καθημερινή ζωή αποκτά πλέον εθνικές διαστάσεις. Παράλληλα δύο εξαιρετικής σημασίας έργα εκπονούνται με στόχο την ενίσχυση της δημιουργίας μίας Εθνικής Υποδομής Χωρικών (Γεωγραφικών) Δεδομένων από το Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης και το Εθνικό Μετσόβειο Πολυτεχνείο με σκοπό να σχεδιάσουν και να προτυποποιήσουν τις οριζόντιες ενέργειες που απαιτούνται για την προσαρμογή της χώρας στο πλαίσιο μίας κοινής δράσης για όλη την Ευρώπη που λίγο αργότερα θα έκανε την εμφάνισή της μέσω του INSPIRE.

Την επόμενη τριετία (2004 - 2006) τα έργα της γεωπληροφορικής και η ευρεία ανάπτυξη διαδικτυακών εφαρμογών παραγωγής, διαχείρισης και διάχυσης της γεωγραφικής πληροφορίας, είναι πλέον ένα τεχνολογικό γεγονός της καθημερινότητας. Με την εξαίρεση του Εθνικού Κτηματολογίου που έχει προσωρινά αναστείλει τις εργασίες του, οι υπόλοιποι φορείς παράγουν νέες προκηρύξεις μεγάλων έργων γεωπληροφορικής. (Το Υπουργείο Αγροτικής Ανάπτυξης και Τροφίμων (πρώην Γεωργίας), συνεχίζει να αναπτύσσει το Ο.Σ.Δ.Ε, το Υπουργείο Ανάπτυξης προκηρύσσει το μεγάλης σπουδαιότητας έργο για τη διαχείριση των υδατικών πόρων της χώρας και η Κοινωνία της Πληροφορίας εξακολουθεί να προκηρύσσει τα έργα της δράσης «Περιφερειακά Γεωγραφικά Συστήματα και Καινοτόμες Ενέργειες» και τα έργα των οργανισμών παραγωγής χαρτογραφικής πληροφορίας έχουν ήδη ξεκινήσει).

Η μεγαλύτερη όμως πρόκληση την περίοδο αυτή, ήταν η έναρξη των συζητήσεων περί της κοινοτικής δράσης "INSPIRE", μίας οδηγίας της Ε.Ε, που συμβάλλει στη θέσπιση κοινών προτύπων για τη γεωγραφική πληροφορία σε όλα τα ευρωπαϊκά κράτη και σε κάθε επίπεδο χρήσης της (παραγωγή, διαχείριση, διάθεση, εκμετάλλευση, κλπ). Η επίσημη ισχύς της ξεκίνησε το 2004 και η ολοκλήρωσή της υλοποίησής της αναμένεται το 2013.

Την τρέχουσα περίοδο, το INSPIRE παίρνει οριστικά σάρκα και οστά μέσω της εφαρμογής του Ν. 3882/2010 «Εθνική Υποδομή Γεωχωρικών Πληροφοριών - Εναρμόνιση με την οδηγία 2007/2/ΕΚ του Ευρωπαϊκού Κοινοβουλίου και του Συμβουλίου της 14ης Μαρτίου 2007» και επίσημα εθνικά γεωγραφικά δεδομένα διατίθενται ανοικτά πλέον από κρατικές ιστοσελίδες (www.opendata.gov.gr, www.ktimatologio.gr κλπ)

Κάθε εποχή παρουσιάζει κάποιες υπερβολές, ιδίως μετά από κάποιες σημαντικές καινοτόμες προσπάθειες και ανακαλύψεις. Είναι γεγονός, πάντως, ότι η ταχύτητα ανάπτυξης των ΣΓΠ, σε σχέση με τα υπόλοιπα συστήματα διαχείρισης πληροφοριών, είναι μικρότερη, λόγω της πολυπλοκότητας των διαδικασιών διαχείρισης των χωρικών δεδομένων, από το βασικό έως τα πλέον προχωρημένα στάδια επεξεργασίας. Αυτό οφείλεται ακριβώς στην πολύπλοκη φύση των χωρικών δεδομένων, αλλά και στις υπάρχουσες μεθοδολογίες ανάκτησης και ανάλυσης των σχετικών πληροφοριών. Πάρα πολλές προκλήσεις στο χώρο των φυσικών και κοινωνικών επιστημών, σε ζητήματα με εγγενή πολυπλοκότητα και διεπιστημονική φύση, παρουσιάζονται ταυτόχρονα με έντονα προβλήματα τα οποία επιζητούν άμεσες λύσεις, όπως για παράδειγμα: «πώς θα αξιολογηθεί η ευπάθεια περιοχών και των πληθυσμών τους στις παγκόσμιες περιβαλλοντικές αλλαγές, πώς θα μετρηθεί και θα υποστηριχθεί η βιοποικιλότητα, πώς θα προβλεφθεί και θα αντιμετωπιστεί το ζήτημα των παγκόσμιων και περιφερειακών επιπτώσεων των ασθενειών, πώς θα πρέπει να γίνει αποτελεσματικότερη διαχείριση της αυξανόμενης κυκλοφορικής ροής των αστικών συγκροτημάτων» κ.ο.κ. Η γεωαναφορά παρέχει έναν θεμελιώδη μηχανισμό σύνδεσης διαφορετικών μορφών δεδομένων, τα οποία είναι απαραίτητα για την επίλυση των προβλημάτων αυτών.

Παρ' όλα αυτά, δεν υπάρχουν πλήρως αναπτυγμένα μοντέλα επίλυσης των πολύπλοκων περιβαλλοντικών και ανθρώπινων συστημάτων που να εκμεταλλεύονται τα «ολοκληρωμένα» αυτά δεδομένα έως την τελική λύση του προβλήματος. Τα γεωγραφικά δεδομένα στην τυπική μορφή τους αφορούν σε αλληλοσυνδεδεμένα και αλληλοεξαρτούμενα φαινόμενα τα οποία ενέχουν κάποια γεωγραφική αναφορά περισσότερο ή λιγότερο σαφή (π.χ. αστικό δίκτυο, οδικό δίκτυο, διοικητικές περιφέρειες) συσχετιζόμενα με στοιχεία μικρότερης χωρικής σαφήνειας (π.χ. ένα οικοσύστημα ή η διασπορά μιας ασθένειας). Η δομή των φυσικών γεωγραφικών στοιχείων και των γεωγραφικών φαινομένων περιγράφεται μέσα στα συστήματα με αντίστοιχες

γεωμετρικές δομές (σημείων, γραμμών, επιφανειών και όγκων) και συσχετίζεται με στατιστικά δεδομένα τα οποία περιγράφουν τις ιδιότητες. Οι γεωγραφικές και χαρτογραφικές αναπαραστάσεις με τη δυναμική υποστήριξη των μεθόδων των ΣΓΠ παρέχουν τη δυνατότητα κάλυψης του κενού των «έτοιμων» μοντέλων επίλυσης χωρικών προβλημάτων, μέσω της διαδοχικής εστίασης σε πολύπλοκες δομές χωρικών φαινομένων διαδικασιών και σχέσεων, κατά τα διαφορετικά στάδια εξέλιξης τους.

4.2 Θεμελιώδεις έννοιες & περιεχόμενο των Σ.Γ.Π. ως ΠΣ.

4.2.1 Ορισμοί - Θεμελιώδη Χαρακτηριστικά

Οι περισσότερες απόπειρες ορισμού των Σ.Γ.Π βασίζονται στην οπτική γωνία από την οποία προσεγγίζεται η έννοια του πληροφοριακού συστήματος, ενώ υπάρχει μεγάλος αριθμός διαφορετικών θεωρήσεων σχετικά με τον προσδιορισμό των συνιστωσών ενός κλασικού θεωρητικού υποδείγματος, για τη διαχείριση γεωγραφικών πληροφοριών. Σε αντίστοιχο αριθμό πονημάτων, σε επίπεδο βιβλιογραφίας και δημοσιεύσεων, απαντώνται διάφορες εκδοχές σχετικά με τον ορισμό και τον αριθμό των συνιστωσών της εφαρμοσμένης εκδοχής των υποδειγμάτων αυτών. Οι τάση η οποία, ίσως καταχρηστικά είχε επικρατήσει για αρκετά χρόνια, είναι εκείνη της προσαρμογής των συνθετικών στοιχείων του συστήματος, ανάλογα με την εφαρμογή (task orientation systems definition).

Δύο ενδιαφέρουσες προτάσεις σε σχέση με τον ορισμό των ΣΓΠ είναι εκείνη της Διεθνούς Ένωσης Γεωδαιτών (Federation Internationale des Geometres - F.I.G., 1983), σύμφωνα με τον οποίο «είναι ένα εργαλείο για λήψη αποφάσεων νομικής, διοικητικής και οικονομικής υφής και ένα όργανο για το σχεδιάσμά και την ανάπτυξη, το οποίο αποτελείται αφ' ενός από μία βάση δεδομένων χωρικού προσδιορισμού (ποιοτικού και ποσοτικού) στοιχείων του γεωγραφικού χώρου και αφ' ετέρου από διαδικασίες και τεχνικές για την συστηματική συλλογή, ενημέρωση, επεξεργασία και διανομή των στοιχείων αυτών» και του Goodchild, σύμφωνα με την οποία τα Σ.Γ.Π. οριοθετούνται ως «<μία οργανωμένη συλλογή μηχανικών υπολογιστικών συστημάτων (hardware), λογισμικών συστημάτων (software), γεωγραφικών δεδομένων και ανθρώπινου δυναμικού, με σκοπό τη συλλογή, καταχώρηση, ενημέρωση, διαχείριση, ανάλυση και απόδοση, κάθε μορφής πληροφορίας που αφορά στο γεωγραφικό περιβάλλον». Ωστόσο, στους ορισμούς αυτούς θα πρέπει να προστεθούν τρία σύγχρονα αιτήματα:

Το αίτημα της «κοινωνικής διάστασης» στην πρόταση περί «νομικής, διοικητικής και οικονομικής υφής» υπό την αμφίδρομη έννοια της σχέσης ανάμεσα στις συνθήκες λήψης αποφάσεων και τις κοινωνικές συνθήκες, όπως π.χ. οι αντιλήψεις που υπάρχουν ανά χώρα για τα δικαιώματα εξουσίας επί του γήινου φυσικού ή ανθρωπογενούς διαθεσίμου, οι πολιτικές που επιλέγονται στους κρίσιμους τομείς της χωροταξίας, της προστασίας περιβάλλοντος, της πολεοδομίας κ.λ.π.:

Το αίτημα για «παρακολούθηση και προστασία του περιβάλλοντος» στο «σχεδιάσμά και την ανάπτυξη» στις σημερινές συνθήκες διαρκούς και έντονης υποβάθμισης, όπου τα συστήματα διαχείρισης πληροφοριών θα πρέπει να κατέχουν σημαντικό ρόλο λόγω της ποσότητας και της πολυπλοκότητας πληροφοριών και προβλημάτων

Το αίτημα της επέκτασης του ορισμού των ΣΓΠ πέραν της σύνθεσης: «εξοπλισμού» «αυτοματοποίησης» - «λογισμικών προγραμμάτων» σε μία σύνθεση που θα συμπεριλαμβάνει με σαφήνεια τον παράγοντα μιας περισσότερο οργανωμένης διαδικασίας αλληλοτροφοδότησης «συστημάτων» «κοινωνικού περιβάλλοντος» και «κέντρων άσκησης χωρικής πολιτικής»

Εκτός από το ονομαστικό "G.I.S." υπάρχει μία σειρά άλλων ονομάτων που μπορούμε να συναντήσουμε διεθνώς, ανάλογα με την εξειδίκευση, τη χρήση ή το επίπεδο ανάλυσης του κάθε συστήματος, αλλά με την ίδια λογική λειτουργίας. Παράδειγμα, τα «Map Information Management Systems» (MIMS) που δίνουν έμφαση στη χαρτογραφία, ή τα «Land Information Systems» (LIS), που απαιτούν κυρίως μεγάλη γεωμετρική ακρίβεια στην απεικόνιση της πληροφορίας και χρησιμοποιούνται σε εφαρμογές «λεπτομερούς» επιπέδου, όπως το κτηματολόγιο, εφαρμογές σε επίπεδο δήμου ή κοινότητας κ.λ.π.

Όσον αφορά στα βασικά χαρακτηριστικά των ΣΓΠ, πέραν των όσων έχουν αναλυθεί στις προηγούμενες παραγράφους, το ενδιαφέρον επικεντρώνεται στις διαφοροποιήσεις που υφίστανται σε σχέση με τα λοιπά πληροφοριακά συστήματα και αφορά στη λειτουργία των

χωρικών μοντέλων που απαιτούνται αλλά και στη δόμηση των στοιχείων εντός αυτών:

- Το εννοιολογικό μοντέλο, που σχετίζεται με τη θεώρηση της γεωγραφικής πραγματικότητας για τον υπό μελέτη χώρο
- Το αναλογικό μοντέλο, που σχετίζεται επίσης με ζητήματα θεώρησης της γεωγραφικής πραγματικότητας σε επίπεδο αφαίρεσης και χαρτογραφικής γενίκευσης των δεδομένων)
- Το χωρικό μοντέλο δεδομένων, (σχηματοποίηση του αναλογικού μοντέλου, σε πρώτη φάση χωρίς περιορισμούς και συμβάσεις που άπτονται της ψηφιακής φύσης των στοιχείων που θα εμπλουτίσουν, εν τέλει, το πληροφοριακό σύστημα)
- Το μοντέλο βάσης δεδομένων, που αφορά στη δομή των δεδομένων σε επίπεδο καταχώρισης στο υπολογιστικό περιβάλλον
- Το υπολογιστικό μοντέλο, που αντιστοιχεί στην καταχώριση και δόμηση των αρχείων στα αποθηκευτικά μέσα του υπολογιστικού συστήματος
- Το μοντέλο διαχείρισης, το οποίο ενσωματώνει παραδοχές και κανόνες λειτουργίας και συσχέτισης του συνόλου των αρχείων που συνθέτει τη βάση δεδομένων.
- Το χαρτογραφικό μοντέλο, στο οποίο ενσωματώνονται κανόνες και διαδικασίες απεικόνισης των δεδομένων.



Σχήμα 4.1 Διάρθρωση λειτουργίας των επί μέρους μοντέλων ενός τυπικού γεωγραφικού πληροφοριακού συστήματος

4.2.2. Συστατικά Μέρη

Υπάρχουν τρία αλληλοτροφοδοτούμενα τυπικά συστατικά μέρη στα παραδοσιακά και σύγχρονα ΣΓΠ που αφορούν: α) μηχανολογικό εξοπλισμό (hard-ware), β) λογισμικές συλλογές προσανατολισμένες σε μια σειρά εφαρμογών (soft-ware modules) και το κατάλληλο οργανωτικό πλαίσιο, συμπεριλαμβανομένων των πληροφοριακών διαθεσίμων (resource-ware) και του εξειδικευμένου επιστημονικού και τεχνικού προσωπικού.

Όσον αφορά στα μηχανικά μέρη, επιγραμματικά αναφέρονται οι τρεις κύριες συνιστώσες: α) η κεντρική μονάδα (επεξεργαστής, αποθηκευτικό μέσο, μνήμες, λειτουργικό σύστημα) β) οι περιφερειακές συσκευές (σαρωτές, ψηφιοποιητές, εκτυπωτικές συσκευές, αποθηκευτικά μέσα κ.λ.π.) και γ) το δίκτυο επικοινωνίας (σύνδεση υπολογιστικών συστημάτων, διαδίκτυο κ.λ.π.).

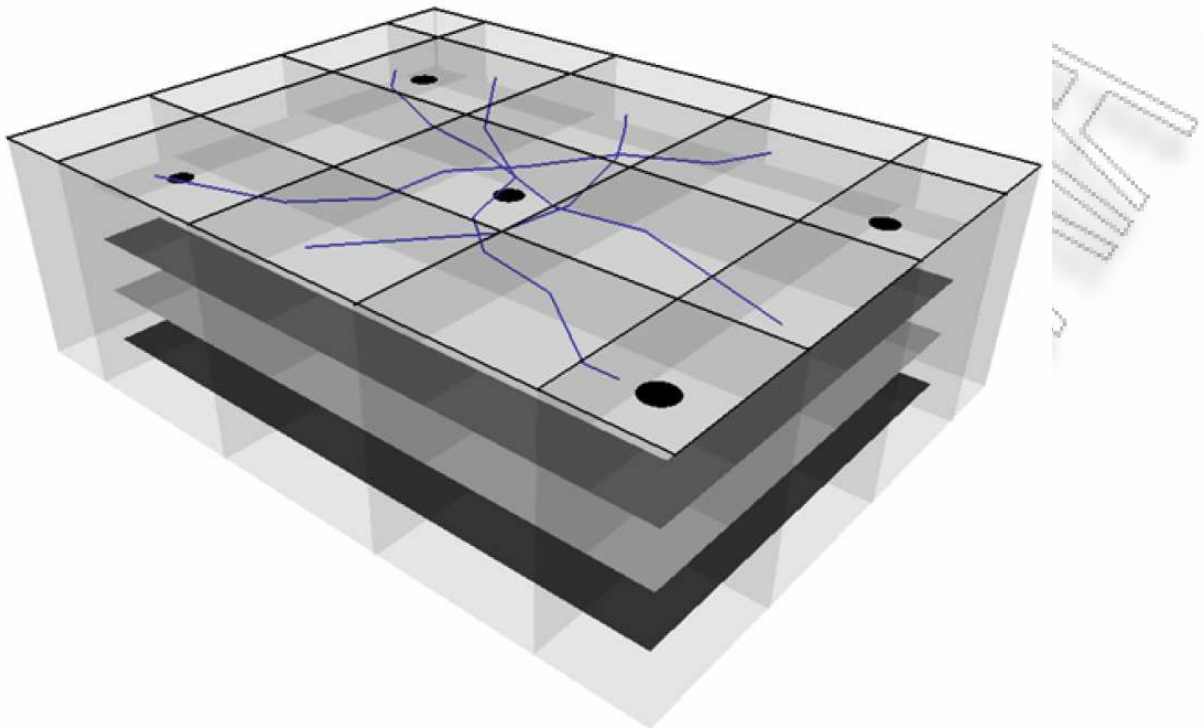
Οι λογισμικές συλλογές ποικίλουν ανάλογα με το συνολικό σχεδιασμό του συστήματος και ταξινομούνται κατά τις έξι λειτουργικές ομάδες:

- Αλγόριθμοι εισαγωγής και επικύρωσης των στοιχείων (data input and verification). Καλύπτουν τις ανάγκες ψηφιακής αναπαραγωγής των συλλεγμένων δεδομένων, εκτελώντας ποικίλους μετασχηματισμούς ανάλογα με την πηγή (αναλογικοί χάρτες, αεροφωτογραφίες, δορυφορικές εικόνες και αισθητήρες, παρατηρήσεις πεδίου κ.ο.κ.). Ο έλεγχος των στοιχείων πραγματοποιείται από ειδικά λογισμικά μέσω των οποίων εξασφαλίζεται τυποποίηση και συμβατότητα.
- Αλγόριθμοι αποθήκευσης στοιχείων και διαχείρισης της βάσης δεδομένων (data storage and database management system). Αφορούν στις μεθοδολογίες δόμησης και οργάνωσης των εξής στοιχείων:
 - ο των γεωμετρικών δεδομένων της γεωγραφικής βάσης (σημεία γραμμές πολύγωνα, ή άλλες πιο περίπλοκες οντότητες αναπαράστασης των γεωγραφικών αντικειμένων)
 - ο των συνδέσεων μεταξύ των γεωμετρικών στοιχείων (τοπολογική δόμηση)
 - ο των περιγραφικών δεδομένων της βάσης (στατιστικά στοιχεία πίνακες)
 - ο της λειτουργίας του συστήματος διαχείρισης βάσεων δεδομένων (DBMS)
 - ο του μοντέλου δεδομένων της γεωγραφικής βάσης
- Λογισμικά διασύνδεσης - επερωτήσεων (interaction and query). Αφορούν σε ένα σύνολο συνθηκών «διαλόγου» ανάμεσα στο σύστημα και τον αναλυτή σχετικά με προσχεδιασμένη ή μη, δομή αναζητήσεων. Η δομή αυτή λειτουργεί είτε σε χωρικό είτε σε περιγραφικό / στατιστικό επίπεδο, είτε σε συνδυασμό τους. Ταυτόχρονα διασφαλίζεται διαδικασία «διεπαφής» (interfaces) προσαρμοσμένης στην ταυτότητα της εφαρμογής.
- Αλγόριθμοι μετασχηματισμών των δεδομένων (data transformation). Συσχετίζονται με διαδικασίες όπως η διόρθωση λαθών, η προετοιμασία των δεδομένων για ενημέρωση των βάσεων, τοπολογικοί μετασχηματισμοί, αποκατάσταση συνθηκών συμβατότητας ανάμεσα στα στοιχεία (π.χ. χαρτογραφικοί μετασχηματισμοί, προβολικών συστημάτων κ.ο.κ.).
- Λογισμικά εξόδου και χαρτογραφικής αναπαράστασης (data output and presentation) που αφορούν στη λειτουργία του χαρτογραφικού μοντέλου και την εξασφάλιση διαδικασιών αποτύπωσης των αποτελεσμάτων σε διάφορα μέσα (χαρτογραφικές συνθέσεις ψηφιακές και αναλογικές, γραφικά αρχεία, αρχεία βίντεο ήχου κ.ο.κ.)

4.3 Οργάνωση - Δόμηση

Οι επικρατούσες «σχολές» οργάνωσης των μοντέλων ΣΓΠ σχετικά την αναπαράσταση της γεωγραφικής πραγματικότητας και με το είδος δόμησης των δεδομένων είναι δύο:

α) Της «συμβατικής ιεραρχίας» (μοντέλο κατά Tomlin, 1990) - σύμφωνα με το οποίο η πληροφορία οργανώνεται σε θεματικό επίπεδα. Τα γεωγραφικά δεδομένα συντίθενται σε μία ιεραρχική δομή που αποτελείται από ένα κεντρικό δυναμικό χαρτογραφικό μοντέλο το οποίο συντίθεται από επί τα μέρους θεματικά επίπεδα. Τα επίπεδα αυτό ορίζονται ως υποσύνολα ομοιογενούς πληροφορίας. Τα δε κριτήρια ομοιογένειας υπόκεινται διπλό παραμετρικό έλεγχο: αφ' ενός σε παραμέτρους εξαρτημένες από το εν λειτουργία μοντέλο (διανυσματικό ή ψηφιδωτό), όπως π.χ. βάσει της διάκρισης «σημειακά, γραμμικά, επιφανειακά δεδομένα» και αφ' ετέρου σε παράγοντες υποκειμενικής φύσης που συναρτώνται με τις ιδιαιτερότητες των εφαρμογών και το συνολικό σχεδιάσμά της γεωγραφικής βάσης.



Σχήμα 4-2 Οργάνωση χαρτογραφικού μοντέλου σε Θεματικά επίπεδα πληροφορίας και «αγκίστρωση» χαρτογραφικού κανάβου

Επίσης, το κεντρικό μοντέλο ελέγχεται γεωμετρικά με διαδικασίες χαρτογραφικών μετασχηματισμών, ώστε τα χρησιμοποιούμενα θεματικά επίπεδα να ευρίσκονται σε κατακόρυφη σύμπτωση βάσει κάποιου χαρτογραφικού προβολικού συστήματος. Το σημαντικότερο πλεονέκτημα του μοντέλου είναι η ανεξαρτησία που παρέχει σε σχέση με το εκάστοτε υιοθετούμενο μοντέλο αναπαράστασης της πραγματικότητας, χρησιμοποιώντας παγίως δύο στοιχειώδεις έννοιες: τη «θέση» (χ,γ,ζ) μέσω της οποίας ορίζονται τόσο τα σημειακά δεδομένα, όσο και τα γραμμικά (ως διαδοχή θέσεων) και την «επιφάνεια» (πολύγωνο, ως μία κλειστή διαδοχή θέσεων).

β) Της «αντικειμενοστραφούς ιεραρχίας» (Object Oriented Hierarchy) Το «αντικείμενο» αποτελεί τη βασική έννοια για τα συστήματά αυτά. Ειδικότερα στις γεωγραφικές βάσεις, κάθε χωρική οντότητα θεωρείται ως ένα αντικείμενο (π.χ. εδαφικές ενότητες, ιδιοκτησίες, κτίσματα, συγκοινωνιακές συνδέσεις και κόμβοι, υδρολογικά ρεύματα, αγωγοί κ.λ.π.). Μία ομοιογενής συλλογή αντικειμένων συνθέτει την «τάξη» αντικειμένων (object class). Σε επίπεδο εφαρμογής, ως κεντρική έννοια θεωρείται η «τάξη» και ο συνολικός σχεδιασμός βασίζεται στον καθορισμό των τάξεων και στον προσδιορισμό σχέσεων ανάμεσα σε αυτές. Αναλυτικότερη περιγραφή του αντικειμενοστραφούς μοντέλου γίνεται πιο κάτω.

4.4 Χρήση - Τομείς Εφαρμογών

Όσον αφορά στην δυναμική της ανάπτυξης των ΓΣΠ στο πεδίο των κοινωνικο-οικονομικών εφαρμογών στην Αστική και Περιφερειακή Ανάπτυξη τα ΣΓΠ αποτελούν βασικό εργαλείο στον Προγραμματισμό και στις διαδικασίες λήψης αποφάσεων. Ειδικότερα στην Ελλάδα, τα Γεωγραφικά Συστήματα Πληροφοριών συμβάλλουν ήδη ή έχουν τη δυναμική να συμβάλλουν ως ολοκληρωμένα εργαλεία στα εξής πεδία:

- Χωροταξικός και Περιφερειακός Προγραμματισμός - Σχεδιασμός (χωρική ανάλυση περιφερειακών ανισοτήτων, διαχείριση ολοκληρωμένων αναπτυξιακών προγραμμάτων και βάσεων κοινωνικό-οικονομικών δεδομένων, επενδυτικά σχέδια και εναλλακτικές στρατηγικές, χωροθετήσεις κατανομών οικονομικών δραστηριοτήτων, αξιολόγηση περιφερειακών και τοπικών αναπτυξιακών προγραμμάτων, συστήματα λήψης αποφάσεων).

Ενδιαφερόμενοι φορείς: Περιφερειακές και Νομαρχιακές Υπηρεσίες, Επιμελητήρια, Ενώσεις, Αναπτυξιακές Εταιρείες κ.ά.

- Αστικός Προγραμματισμός - Σχεδιασμός (χωρική ανάλυση αστικών περιοχών, δήμων, γειτονιών-ανισότητες, διαχείριση ολοκληρωμένων προγραμμάτων αστικής ανάπτυξης, πολιτική αναπλάσεων, πολιτική χρήσεων γης, δόμηση, κτηματολόγιο). Ενδιαφερόμενοι φορείς: Δημόσιοι φορείς, ΥΠΕΧΩΔΕ, Τοπική Αυτοδιοίκηση κ.ά.
- Περιβάλλον (Διαχείριση οικοσυστημάτων, πολιτικής προστασίας και πρόληψης, συστήματα λήψης αποφάσεων και εκτίμηση επιπτώσεων, υποδείγματα αλληλεπιδράσεων οικονομικών και περιβαλλοντικών συστημάτων, διαχείριση υδατικών πόρων, επιχειρησιακή έρευνα). Ενδιαφερόμενοι φορείς: ΥΠΕΚΑ, ΠΕΡΠΑ, Τοπική Αυτοδιοίκηση.
- Κτηματολόγιο - Καταγραφή διαχείριση δημόσιων και ιδιωτικών ακινήτων Ενδιαφερόμενοι φορείς: Δημόσιοι και ιδιωτικοί φορείς, Τοπική Αυτοδιοίκηση κ.ά.
- Αγροτική ανάπτυξη και πολιτική (χρήσεις γης, ταξινόμησης, παραγωγικότητα γεωργικής γης, επιδοτήσεις κ.ά. Ενδιαφερόμενοι φορείς: Υπ.Α.Α.Τ, άλλοι δημόσιοι και ιδιωτικοί φορείς, Τοπική Αυτοδιοίκηση κ.ά.
- Πολιτισμός - Τουρισμός (ανάπτυξη βάσεων γεωγραφικών δεδομένων για τους αρχαιολογικούς χώρους, ανασκαφικές δραστηριότητες, χωροθέτηση τουριστικών μονάδων και δημοσίων τουριστικών ακινήτων, τουριστική ανάπτυξη περιοχών κ.ά.) ΥπΠοΤ, άλλοι δημόσιοι και ιδιωτικοί φορείς, Τοπική Αυτοδιοίκηση κ.ά.
- Συγκοινωνίες - Μεταφορές (διαχείριση συστημάτων μεταφορών (οδικών, ακτοπλοϊκών, αεροπορικών), διαχείριση αστικών συγκοινωνιών, πολιτική πρόληψης ατυχημάτων, κ.ά) Ενδιαφερόμενοι φορείς: Φορείς συγκοινωνιών (Δημόσιοι- Ιδιωτικοί), Τοπική Αυτοδιοίκηση κ.ά.
- Τεχνική υποδομή (διαχείριση δικτύων ύδρευσης-αποχέτευσης, ενέργειας, τηλεπικοινωνιών, προσδιορισμός περιοχών εξυπηρέτησης, χωροθετήσεις-κατανομές κ.ά.) Ενδιαφερόμενοι φορείς: Αρμόδιοι δημόσιοι φορείς υποδομών, ΔΕΗ, ΟΤΕ, ΕΥΔΑΠ, ΔΕΦΑ, Τοπική Αυτοδιοίκηση.
- Φορολογία (Φορολογία ακίνητης περιουσίας, διαχείριση φορολογικών στοιχείων). Ενδιαφερόμενοι φορείς: Υπ. Οικονομικών, Οικ. Εφορία, Τοπική Αυτοδιοίκηση.
- Εκπαίδευση και Υγεία - Πρόνοια (πολιτική διαχείρισης παροχών εκπαίδευσης, υγείας-πρόνοιας, περιοχές ειδικών χαρακτηριστικών, χωροθετήσεις-κατανομές κέντρων εξυπηρέτησης, περιοχές εξυπηρέτησης, διασπορά ασθενειών, επιδημιολογικές μελέτες κ.ά.)- Ενδιαφερόμενοι φορείς: Υπουργείο Παιδείας, Υγείας Πρόνοιας, Νομικές. Υπηρεσίες, Τοπική Αυτοδιοίκηση κ.ά.
- Πυροσβεστική, Δασική Υπηρεσία, κ.λ.π. (πολιτικές πρόληψης και αντιμετώπισης εκτάκτων αναγκών, ελαχιστοποίηση διαδρομών, κόστους κ.ά.) Ενδιαφερόμενοι φορείς: Νομαρχίες, Πυροσβεστική Υπηρεσία, Υπ. Γεωργίας - Δασική Υπηρεσία κ.ά.
- Marketing (Ανάλυση καταναλωτικής συμπεριφοράς, συστήματα λήψης αποφάσεων) Ενδιαφερόμενοι φορείς: Ιδιωτικός τομέας.
- Αγορά Εργασίας (χωρική ανάλυση αγορών εργασίας, σύζευξη προσφοράς-ζήτησης, πολιτικές απασχόλησης, ανεργίας και επαγγελματικής κατάρτισης, κινητικότητα εργατικού δυναμικού, μετακινήσεις τόπου εργασίας-κατοικίας) Ενδιαφερόμενοι φορείς: Υπ. Εργασίας, ΟΑΕΔ, Φορείς επαγγελματικής κατάρτισης.
- Δίκτυα διανομών, πωλήσεων και χωροθετήσεις κατανομών (ανάλυση και διαχείριση δικτύων διανομών προϊόντων και υπηρεσιών, βελτιστοποίηση διαδρομών τροφοδοσίας, χωροθετήσεις κέντρων παροχών). Ενδιαφερόμενοι φορείς: Επιχειρήσεις δημόσιου και ιδιωτικού τομέα.

Τα ανωτέρω πεδία εφαρμογών, τα οποία προφανώς δεν ολοκληρώνουν το ευρύ φάσμα εφαρμογών οι οποίες μπορούν να αναπτυχθούν, δείχνουν τις δυνατότητες ανάπτυξης των Συστημάτων Γεωγραφικών Πληροφοριών. Δεδομένης της συνθετότητας των αναπτυξιακών προβλημάτων στις πόλεις και στις περιφέρειες, τα ΣΓΠ μπορούν να συμβάλλουν στην ενιαία καταγραφή, οργάνωση, διαχείριση και ανάλυση των κοινωνικοοικονομικών δεδομένων, ως προϋποθέσεις για τις διαδικασίες λήψης αποφάσεων στην αστική και περιφερειακή ανάπτυξη. Και αυτό αφορά τόσο στον Δημόσιο Τομέα, όσο και στον Ιδιωτικό, ο οποίος μάλιστα σε

συγκεκριμένες περιπτώσεις (πολυεθνικές εταιρείες, μεγάλες επιχειρήσεις του δευτερογενούς και τριτογενούς τομέα) έχει να επιδείξει σημαντικές εφαρμογές των ΣΓΠ στα συγκεκριμένα αντικείμενα του ενδιαφέροντος του (π.χ. δίκτυα παραγωγής, διανομές προϊόντων και υπηρεσιών, χωροθετήσεις).

4.5 Χωρικά προβλήματα και χωρικός σχεδιασμός: Η οπτική των συστημάτων πληροφοριών

4.5.1 Τυπικά χωρικά προβλήματα και ΣΓΠ

Η συστηματική γνώση και η σαφής οριοθέτηση των χωρικών προβλημάτων αποτελεί κλασική μεθοδολογική προαπαιτήτηση στο σχεδιάσμά και τη χρήση χωρικών συστημάτων πληροφοριών, Η ανάλυση των προβλημάτων σε επιστημονικό επίπεδο και η επίλυση τους σε πρακτικό, απαιτεί διερεύνηση των παραμέτρων του χωρικού προβλήματος και ανάδειξη ενός πλήρους συνόλου ειδικών ερωτημάτων που θα ενσωματώνονται στο πληροφοριακό σύστημα. Στη συνέχεια θα μετασχηματίζονται σε δεδομένα και διαδικασίες με τρόπο ώστε να εξασφαλίζεται η αποτελεσματική λειτουργία του. Μία πρώτη ταξινόμηση των ερωτημάτων τα οποία συσχετίζονται με το γεωγραφικό χώρο και συνδέονται με τα τυπικά προβλήματα που αντιμετωπίζονται σε αυτόν, υπό την οπτική των ΣΓΠ είναι:

- Ερωτήματα που συσχετίζονται με το χαρακτήρα των χωρικών οντοτήτων (τον ποσοτικό, ως τιμή των χωρικών μεταβλητών, τον ποιοτικό, ως ειδοποιού διαφοράς μιας οντότητας σε σχέση με άλλες, αλλά και συνδυασμοί των δύο)
- Ερωτήματα που συσχετίζονται με μετρήσεις επί του αριθμού των οντοτήτων που πληρούν συγκεκριμένες συνθήκες (βάσει της τιμής, των ιδιοτήτων ή της γεωγραφικής θέσης τους)
- Ερωτήματα που συσχετίζονται με την απόσταση των οντοτήτων από συγκεκριμένα σημεία του χώρου ή άλλες χωρικές οντότητες (απλή ευκλείδεια ή υπολογιζόμενη μέσω κάποιας πιο σύνθετης συνάρτησης)
- Ερωτήματα που συσχετίζονται με την ένταξη της οντότητας εντός συγκεκριμένων ζωικών ορίων, καθοριζόμενων είτε από κάποιο γεωμετρική μορφή (π.χ. ένας κύκλος, η μία πιο σύνθετα υπολογιζόμενη γεωμετρική επιφάνεια) είτε από κάποια άλλη γεωγραφική ή ανθρωπογενή οριοθέτηση (π.χ μία ήπειρος ή μία χώρα)

Τα χωρικά προβλήματα εντασσόμενα με τη σειρά τους στα γεωγραφικά πληροφοριακά συστήματα, συσχετίζονται:

με μετρήσεις επί της γήινης επιφάνειας,

με χωρικές σχέσεις που αναπτύσσονται ανάμεσα στις οντότητες,

με τη διερεύνηση τρόπων έμμεσου προσδιορισμού ιδιοτήτων για αντικείμενα που δεν είναι άμεσα παρατηρήσιμα (π.χ. με τη χρήση δορυφορικής τεχνολογίας)

με τη δημιουργία κατάλληλων υποδειγμάτων ανάλυσης και πρόβλεψης, σύμφωνα με τη φύση και τους επιστημονικούς τομείς στους οποίους τα προβλήματα μπορούν να υπαχθούν.

Στο επίπεδο της χωρικής ανάλυσης η οποία πραγματοποιείται κατά τη διαδικασία οριοθέτησης και επίλυσης χωρικών προβλημάτων, τα γεωγραφικά πληροφοριακά συστήματα απαντούν σε ερωτήματα τα οποία ταξινομούνται σε πέντε τυπικές κατηγορίες:

Τοπογραφία: «Τι βρίσκεται πού ...» Η ουσία της ερώτησης βρίσκεται στο γεγονός ότι απαιτείται ακριβή γνώση του τι υπάρχει σε κάθε συγκεκριμένο τόπο του χώρου μελέτης. Η ταυτότητα ενός τόπου μπορεί να περιγράψει με διάφορους τρόπους, όπως για παράδειγμα με το όνομα της τοποθεσίας, κάποιο γεωγραφικό κωδικό, ή με κάποιο γραφικό συμβολισμό σε συνδυασμό με ένα σύστημα γεωγραφικών συντεταγμένων όπως το γεωγραφικό μήκος και πλάτος, συστήματα καρτεσιανών συντεταγμένων κ.λ.π.

Βάσει συνθήκης: «Που βρίσκεται...» Η δεύτερη αυτή ερώτηση είναι τρόπον τινά η αντιστροφή της πρώτης και απαιτεί στοιχεία χωρικής ανάλυσης για να απαντηθεί. Αντί για τον προσδιορισμό της ταυτότητας ενός δεδομένου τόπου, ζητείται να βρεθεί ο γεωμετρικός τόπος μέσα στον οποίο ικανοποιούνται ορισμένες συνθήκες, (π.χ. πού βρίσκεται γήινη επιφάνεια χαρακτηρισμένη δάσος και με εμβαδόν μεγαλύτερο των 100 στρεμμάτων, να απέχει απόσταση

μικρότερη των 60 χιλιομέτρων από το αστικό κέντρο και 2 απόσταση χιλιομέτρων από το οδικό δίκτυο ... κ.ο.κ.)

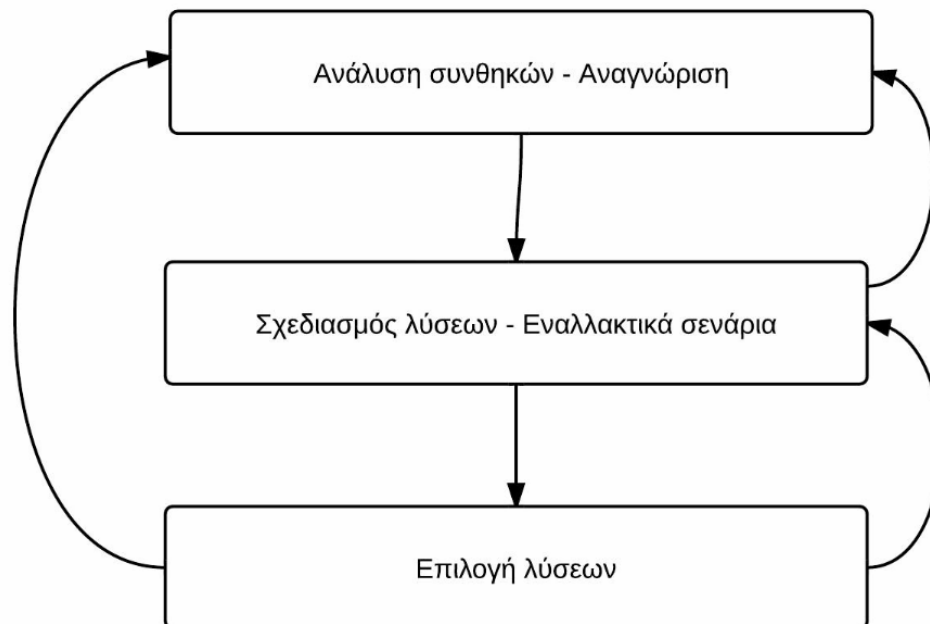
Τάσεις: «Ποια η μεταβολή ...» Η ερώτηση αυτή προαπαιτεί την απάντηση των δύο προηγούμενων αναζητώντας διαφορές που ανακύπτουν κατά την πάροδο του χρόνου. Στόχο αποτελεί η δημιουργία προτύπων και υποδειγμάτων ανάλυσης σχετικών με τις μεταβολές στο χώρο, τα οποία αποτελούν τη βάση για κάθε απόπειρα μελλοντικών προβλέψεων.

Πρότυπα: «Από ποια χωρικά πρότυπα χαρακτηρίζεται...» Στις διαδικασίες ανάδειξης κανόνων συμπεριφοράς και προτύπων των χωρικών φαινομένων αναζητούνται συσχετισμοί και νόμοι οι οποίοι διέπουν φαινόμενα που συμβαίνουν ταυτόχρονα, (ή είναι το ένα συνέπεια του άλλου) και αφορούν σε ένα συγκεκριμένο χώρο.

Υποδείγματα: «Τι θα συνέβαινε αν...» Η απάντηση αυτού του είδους ερωτήσεων απαιτεί συνολικότερη επιστημονική θεώρηση υπό τη έννοια ότι μόνη η γεωγραφική πληροφορία ή οι όποιες διαδικασίες μέσω ΣΓΠ δεν επαρκούν ως παράμετρος ανάλυσης πολύπλοκων, φαινομένων. Η συνήθης μεθοδολογία βασίζεται σε μοντέλα πρόβλεψης και προσομοίωσης λειτουργιών (predictive spatial models / simulation models). Παράδειγμα: τι θα συνέβαινε εάν ένας καινούριος δρόμος ενταχθεί στο οδικό δίκτυο, ή ποιες θα ήταν οι επιπτώσεις μιας πιθανής μόλυνσης του υδροφόρου ορίζοντα μιας περιοχής... κ.ο.κ.

4.5.2 ΣΓΠ και λήψη αποφάσεων

Έχει ήδη αναφερθεί, πως ο σημαντικότερος, ίσως, στόχος της χρήσης γεωγραφικών συστημάτων πληροφοριών είναι η παροχή υποστήριξης κατά τις διαδικασίες λήψης χωρικών αποφάσεων. Για την επίτευξη του στόχου αυτού απαιτείται επεξεργασμένη πληροφορία η οποία υπόκειται σε προσαρμογές σχετικές προς τις ανάγκες των αντίστοιχων φορέων λήψης αποφάσεων. Οι προσαρμογές αυτές υλοποιούνται κατ' αρχάς μέσω του μετασχηματισμού των χωρικών δεδομένων σε κατάλληλη πληροφορία μέσω της οποίας θα περιγραφεται η υπάρχουσα κατάσταση αλλά και τα σενάρια επίλυσης ενός χωρικού προβλήματος.



Σχήμα 4-3 Τυπικός σχεδιασμός της διαδικασίας λήψης αποφάσεων σε τρία στάδια.

Βεβαίως κατά τη διαδικασία αυτή, υπάρχει πληθώρα μεθοδολογιών προσέγγισης και πλαισίων ανάλυσης. Μία από τις ευρύτερα αποδεκτές μεθοδολογίες, έχει προταθεί από τον Simon, ο οποίος εισηγήθηκε μία δομή προσέγγισης με τρία στάδια ανάλυσης: το στάδιο της αναγνώρισης του προβλήματος, το στάδιο του σχεδιασμού εναλλακτικών σεναρίων-λύσεων και

το στάδιο της επιλογής λύσεων.

Και στα τρία στάδια, η συμβολή των ΓΣΠ μπορεί να είναι σημαντική. Δεν είναι απαραίτητη μία γραμμική ακολουθία διαδρομής για τα τρία αυτά στάδια, αλλά είναι δυνατόν κάποιο στάδιο να επαναλαμβάνεται πολλές φορές, έως την τελική επιλογή λύσεων, πράγμα το οποίο απαιτεί ευελιξία και ταχύτητα διαδικασιών. Κάθε στάδιο απαιτεί διαφορετικούς τύπους πληροφορίας και διαφοροποιημένες διαδικασίες ανάλυσης ενώ το κρίσιμο ερώτημα είναι το πώς και σε ποια έκταση τα γεωγραφικά πληροφοριών μπορούν να παράσχουν υποστήριξη (κατάλληλη πληροφορία) σε κάθε ένα από τα τρία στάδια.

4.5.2.1 Ανάλυση συνθηκών – Αναγνώριση προβλημάτων

Κάθε διαδικασία αποφάσεων χωρικής υφής, έχει ως αφετηρία την αναγνώριση, τον ακριβή προσδιορισμό, μίας σειράς χωρικών προβλημάτων. Σε γενικευμένη προσέγγιση, ένα χωρικό πρόβλημα μπορεί να ορισθεί ως η απόσταση, το κενό, ανάμεσα στην υπάρχουσα και την επιθυμητή κατάσταση ενός γεωγραφικού συστήματος (σύμφωνα με τον τρόπο κατά τον οποίο αμφότερες, γίνονται αντιληπτές από τον αναλυτή). Οι αποφάσεις είναι αναγκαίες, όταν λόγω περιστάσεων ευνοείται κάποια παρέμβαση, όταν ένα ή περισσότερα από τα προβλήματα παρουσιάζει ιδιαίτερη όξυνση, όταν μία κατάσταση αποκλίνει σημαντικά από τον αρχικό σχεδιασμό, ή όταν κάποια σημαντικά συνθετικά στοιχεία της χωρικής πραγματικότητας επιδέχεται βελτίωσης.

Κατά συνέπεια, η φάση της αναγνώρισης εμπεριέχει έρευνα, ανίχνευση και αξιολόγηση συνθηκών και οριοθέτηση αναγκών οι οποίες απαιτούν παρέμβαση. Ταυτόχρονα, τα αποτελέσματα της ερευνητικής αυτής διαδικασίας θα πρέπει να συστηματοποιούνται, να ιεραρχούνται και να αποτυπώνονται ως στοιχεία μίας οργανωμένης βάσης γεωγραφικών δεδομένων. Τούτο βεβαίως απαιτεί κάποιες παραδοχές και υποθέσεις (ή θεωρήσεις της πραγματικότητας) οι οποίες θα αποτελέσουν τη βάση προσδιορισμού του προβλήματος. Οι υποθέσεις αυτές έχουν σχέση με ένα θεμελιώδες ερώτημα των συστημάτων, το οποίο αφορά στο ποιος θα πρέπει να είναι οι χωρικές οντότητες οι οποίες θα τεθούν υπό παρακολούθηση, πώς θα επιλεγούν, θα ταξινομηθούν και εν τέλει θα εγγραφούν ως δεδομένα. Ιδιαίτερη διερεύνηση θα πρέπει να υπάρξει στην αξιολόγηση τη χρησιμότητα, την ενημερότητα, τον απαιτούμενο βαθμό ακρίβειας, την αξιοπιστία και την προσαρμοστικότητα των δεδομένων, ώστε να αποτελέσουν ένα ολοκληρωμένο σύνολο. Τα δεδομένα που προκύπτουν από τη διερευνητική αυτή ανάλυση, παίζουν τον σημαντικότερο ρόλο τη φάση της αναγνώρισης του προβλήματος, παρά το γεγονός ότι στα ΣΓΠ υπάρχουν και πρόσθετες τεχνικές (γεωγραφικής ανάλυσης κυρίως) οι οποίες με την κατάλληλη προσαρμογή, καθίστανται εξ' ίσου σημαντικές, αναλόγως και προς την ιδιαιτερότητα των προβλημάτων της κάθε περίπτωσης.

Στον επιστημονικό χώρο της γεωγραφικής και χωρικής ανάλυσης γίνεται συζήτηση σχετική με το κατά πόσον - και με αυτοτελή τρόπο - τα εμπορικά συστήματα γεωγραφικής πληροφορίας μπορούν να υποστηρίξουν με επάρκεια τη συγκεκριμένη φάση της αναγνώρισης των προβλημάτων. Έως σήμερα, η αναμενόμενη από τα συστήματα υποστήριξη εστιάζεται σε προβλήματα τα οποία είναι παραδοσιακά συνδεδεμένα με τη συλλογή και ανάλυση των δεδομένων, όπου η συμβολή τους είναι ουσιαστική και αποτελεσματική. Ο ρόλος των συστημάτων στη αρχική αυτή φάση της διαδικασίας λήψης αποφάσεων, προσδιορίζεται στη συσσώρευση και διαχείριση μεγάλων ποσοτήτων δεδομένων και πληροφοριών και την ολοκλήρωση δεδομένων τα οποία μπορεί να προέρχονται από σειρά πολλών και ετερόκλητων πηγών. Για παράδειγμα, τα υδρολογικά δεδομένα που συσχετίζονται με την καταλληλότητα του πόσιμου ύδατος είναι δυνατόν να συνδυαστούν με τα επιδημιολογικά και υγειονομικά δεδομένα μίας περιφέρειας, προκειμένου να προσδιοριστεί ο συσχετισμός του πόσιμου ύδατος με μία ασθένεια, η κατανομή της στο χώρο αλλά γενικότερα να υποβοηθηθεί υγειονομική πολιτική στη συγκεκριμένη περιοχή. Τα σύγχρονα ΓΣΠ παρέχουν συνδυαστικές λειτουργίες ολοκλήρωσης ετερογενών δεδομένων με στόχο την αποτελεσματική οριοθέτηση ποικίλων χωρικών προβλημάτων τόσο στον ιδιωτικό, όσο και στον δημόσιο τομέα. Σημαντική είναι η συμβολή των συστημάτων/ στην κατανομή και αποτελεσματική παρουσίαση πληροφοριών οι οποίες με τις συμβατικές μεθόδους παρουσίασης απαιτούν χιλιάδες ίσως, σελίδες στοιχείων υπό μορφή πινάκων, διαγράμμάτων και χαρτογραφικών συνθέσεων ή σκαριφημάτων.

4.5.2.2 Σχεδιασμός λύσεων – Εναλλακτικά σενάρια

Η φάση αυτή περιλαμβάνει την ανίχνευση, ανάλυση και πλήρη ανάπτυξη μίας σειράς πιθανών λύσεων για το συνολικό πρόβλημα όπως αυτό οριοθετήθηκε κατά τη φάση της ανάλυσης. Οι εναλλακτικές λύσεις μπορούν να κινούνται στο πλαίσιο διαφόρων εναλλακτικών μορφών δράσης (όπως για παράδειγμα; υιοθέτηση προτύπων χρήσεων γης, αναδιοργάνωση διοικητικών περιφερειών, καταμερισμός φυσικών πόρων σε συγκεκριμένες περιοχές ή δραστηριότητες, κατάργηση / ενίσχυση οικονομικών δραστηριοτήτων ή σε λεπτομερέστερο επίπεδο, κάποιας συγκεκριμένης εγκατάστασης κ.ο.κ.).

Στην τυπική περίπτωση, γίνεται χρήση κάποιου χωρικού υποδείγματος το οποίο βοηθά τον αναλυτή, στον καθορισμό μίας σειράς εναλλακτικών σεναρίων σχετικών με την επίλυση του προβλήματος. Το μοντέλο αυτό αποτελεί πάντοτε μία απλοποιημένη απόδοση ή αφαίρεση της χωρικής πραγματικότητας (με δεδομένο ότι η πολυπλοκότητα της πραγματικότητας δεν επιτρέπει σχεδόν ποτέ, την «ακριβή προσομοίωση» της) και υλοποιείται συνήθως με τη μορφοποίηση αντίστοιχων βάσεων γεωγραφικών δεδομένων (geographic databases). Η γεωγραφική βάση μπορεί να έχει διάφορες μορφές και εκδόσεις στην προσπάθεια αποκατάστασης του κατάλληλου περιβάλλοντος για την ανάδειξη των σεναρίων - λύσεων, μέσα από τις λειτουργίες επεξεργασίας και ανάλυσης δεδομένων και πληροφοριών του GIS.

Οι δυνατότητες που δίδονται μέσω των GIS στη φάση αυτή, οριοθετούνται κυρίως από τις κλασικές αρχές των χωρικών σχέσεων των συστημάτων: σχέσεις σύνδεσης (connectivity), συνέχειας (contiguity), εγγύτητας (proximity) και τις μεθόδους «χαρτογραφικής άλγεβρας» (map algebra - overlay methods). Οι μέθοδοι αυτές χρησιμοποιούνται πολύ συχνά για τη διερεύνηση καταλληλότητας χώρων με σκοπό τη χωροθέτηση δραστηριοτήτων (π.χ. τουριστικές ζώνες, βιομηχανικές περιοχές, σχολεία, νοσοκομείο κ.ο.κ.). Στις τις διαδικασίες αυτές ο ρόλος των ΣΓΠ περιορίζεται ουσιαστικά στην επεξεργασία μιας σειράς ψηφιακών θεματικών επιπέδων ομοιογενούς γεωγραφικής πληροφορίας (layers) για τον προσδιορισμό γεωμετρικών τόπων οι οποίοι καλύπτουν ταυτόχρονα ορισμένο προκαθορισμένα κριτήρια της χωροθέτησης.

Κατά τη φάση της δημιουργίας των σεναρίων, ένα σημαντικό μειονέκτημα της μεθοδολογίας αυτής αναδεικνύεται σε περιπτώσεις κατά τις οποίες υπάρχουν ασυμβατότητες ανάμεσα στα κριτήρια. Στη περίπτωση αυτή, η διαδικασία μέσω ΣΓΠ δεν δίνει επαρκή αναλυτική υποστήριξη, στο βαθμό μάλιστα που τα συστήματα δεν παρέχουν μεγάλες δυνατότητες ενσωμάτωσης των υποκειμενικών κριτηρίων του αναλυτή. Επιπλέον, οι πολύπλοκες σχέσεις που αναπτύσσονται σε ορισμένα χωρικά προβλήματα είτε δεν είναι χαρτογραφήσιμες, είτε απαιτούν την κατασκευή δυσκίνητων και χρονοβόρων χαρτογραφικών σχημάτων. Κατά συνέπεια, τα «μοντέλα αποφάσεων» λειτουργούν στο «παρασκήνιο» του περιβάλλοντος των ΣΓΠ, χρησιμοποιώντας το συχνά ως μηχανισμό υποστήριξης διαδικασιών σχετικά αποκομμένων από τη λογική του εμπορικού «πακέτου» που χρησιμοποιείται.

Κατά την έως πρόσφατα χρησιμοποιούμενη πρακτική, τα μειονεκτήματα αυτά καλύπτονται μέσω διαδικασιών δόμησης πρόσθετων λογισμικών διαδικασιών (customization) οι οποίες καλούνται να ενσωματώσουν τις λογικές του αναλυτή στο σύστημα, χρησιμοποιώντας ήδη αναπτυγμένα συστήματα ανάλυσης αποφάσεων (decision analysis systems). Τα τελευταία χρόνια έχουν γίνει προσπάθειες ένταξης ορισμένων αναλυτικών λειτουργιών στα εμπορικά συστήματα, ιδίως στον τομέα της «προσομοίωσης» (simulation) και «βελτιστοποίησης» (optimization) των μοντέλων αποφάσεων. Επίσης, σημαντικά βήματα έχουν γίνει στον τομέα των δυνατοτήτων χωρικής ανάλυσης των συστημάτων γεωγραφικών πληροφοριών, όπως για παράδειγμα οι λογισμικές συλλογές μεθόδων ποσοτικής γεωγραφίας και γεωστατιστικής ανάλυσης (π.χ. χωρική αυτοσυσχέτιση - spatial autocorrelation, χωρικές αλληλεπιδράσεις - spatial interactions, μοντέλα χωροθέτησης- κατανομών - location- location modeling κ.λ.π.). Σημειώνεται πως η εξέλιξη αυτή εντοπίζεται όλο και περισσότερο σε λογισμικές συλλογές με εξειδικευμένο προσανατολισμό όπως για παράδειγμα συστήματα σχεδιασμού μεταφορών, συστήματα διαχείρισης υδάτων κ.ά.

4.5.2.3 Επιλογή σεναρίων – λύσεων

Το προηγούμενο στάδιο αφορά κυρίως στο σχεδιάσμα των εναλλακτικών λύσεων. Το στάδιο της επιλογής αναλύεται σε δύο φάσεις: σε πρώτη φάση γίνεται πλήρης ανάπτυξη των λύσεων αυτών ενώ στη δεύτερη φάση επιχειρείται αποκατάσταση υποθετικών συνθηκών

λειτουργίας, με τη μορφή σεναρίων που συμπεριλαμβάνουν όλες τις εκδοχές επίλυσης. Κάθε σενάριο εξετάζεται συνδυαστικά με άλλα και στο πλαίσιο προκαθορισμένων κανόνων επί των ενδεχομένων αποφάσεων και των αντικειμενικών στόχων που τις καθορίζουν. Οι κανόνες αυτοί είναι απαραίτητοι για την ταξινόμηση και την ιεράρχηση των σεναρίων που αποτελούν και την κλασική μεθοδολογική διαδικασία στη λήψη αποφάσεων. Παραδοσιακά η διαδικασία αυτή επιτελείται εκτός των πλαισίων λειτουργίας του GIS και σύμφωνα με τις ιδιαίτερες μεθοδολογικές προσεγγίσεις του αναλυτή.

Στο στάδιο αυτό, ως κρίσιμο σημείο σχετικά με τη χρήση των GIS, καθορίζεται η προσπάθεια ενσωμάτωσης των μεθοδολογικών προσεγγίσεων εντός της ίδιας της λογικής και τις επεξεργασίες του γεωγραφικού συστήματος. Η συνήθης δυσκολία εδώ, εντοπίζεται στην αδυναμία δυναμικής αξιολόγησης των αλλαγών στα κριτήρια, αλλά και στο βάρος με το οποίο τα ίδια συμμετέχουν μέσα στο σύστημα αποφάσεων. Ο τομέας στον οποίο τα GIS συμβάλλουν αποφασιστικά έως σήμερα, αφορά στην ελαχιστοποίηση των ασυμβατοτήτων επί των δεδομένων και των παρεπόμενων εμπλοκών επί των διαδικασιών επιλογής. Αυτό καθίσταται δυνατό μέσω της παροχής δεδομένων και πληροφοριών κατάλληλης ποσότητας και ποιότητας, συμβάλλοντας στην κατάργηση φαινομένων σύγχυσης που συχνά κυριαρχούν λόγω της πληθώρας και δυσκολιών στην αξιολόγηση μεγάλου όγκου πληροφορίας.

Στα αρχικά στάδια της ανάλυσης των σεναρίων, η διαθεσιμότητα μεγάλου όγκου αξιολογημένης πληροφορίας δρα θετικά. Σε πιο προχωρημένα στάδια και καθώς οι παράγοντες ανάλυσης αυξάνουν, αντιμετωπίζονται ασυμβατότητες υψηλότερου επιπέδου με τη μορφή «ανταγωνιστικότητας των κριτηρίων». Στη φάση αυτή απαιτείται η παρουσία διαχειριστικών συνδυασμών που αφορούν στη γεωγραφική πληροφορία και στην εξειδίκευση των «συστημάτων υποστήριξης αποφάσεων» (Decision Support Systems - DSS και Spatial Decision Support Systems - SDSS).

Υπό τη λογική των τεχνικών ανάλυσης αποφάσεων που υποστηρίζονται από τα ανωτέρω συστήματα, η υποστήριξη και η επιτυχία των GIS στο στάδιο της επιλογής λύσεων, εντοπίζεται στον έλεγχο της ροής και το φιλτράρισμα των πληροφοριών ανάμεσα σε μεγάλες γεωγραφικές βάσεις και στα συνιστώσα (σχετικά με τη παροχή πληροφορίας) μέρη των συστημάτων υποστήριξης αποφάσεων.

4.6 Συνδυασμός τεχνολογιών G.I.S. και OLAP

Εκτιμάται ότι ένα μεγάλο ποσοστό των δεδομένων που χρησιμοποιούνται από τις επιχειρήσεις και άλλους οργανισμούς μπορούν να συσχετιστούν και με μία χωρική-γεωγραφική διάσταση (λόγου χάρη ταχυδρομική διεύθυνση, τόπωνύμιο, γεωγραφικές συντεταγμένες). Η δυνατότητα παρουσίασης δεδομένων πάνω σε χάρτες και σύγκρισης χαρτών διαφορετικών φαινομένων και χρονικών περιόδων σε σχέση με πίνακες και στατιστικά γραφήματα, προσδίδουν μία νέα διάσταση στη χρησιμότητα των χωρικών δεδομένων. Επιπροσθέτως η απαίτηση για γρήγορη χωρική και χρονική ανάλυση δεδομένων, αντλώντας τα πλέον και από χάρτες και όχι μόνο από πίνακες αξιοποιείται από ολοένα και μεγαλύτερο αριθμό χρηστών. Η απαίτηση αυτή μπορεί να ικανοποιηθεί με την συνεκμετάλλευση των δυνατοτήτων των γεωγραφικών πληροφοριακών συστημάτων (GIS) και των πολυδιάστατων βάσεων δεδομένων. Για την περιγραφή της παραπάνω τεχνολογίας χρησιμοποιείται ο όρος Spatial Online Analytical Processing (SOLAP).

5. Υλοποίηση συστήματος αναλυτική επεξεργασία μεταναστευτικών δεδομένων και οπτικοποίησης γεωγραφικής πληροφορίας

Σκοπός του παρόντος κεφαλαίου είναι η περιγραφή των διαδικασιών για την υλοποίηση μιας βάσης δεδομένων, που θα μπορεί να δεχτεί τα στοιχεία των αλλοδαπών ασφαλισμένων του ΟΑΕΕ και η οποία θα έχει σαν ρόλο την παροχή αυτών ώστε να εισαχθούν σε αποθήκη δεδομένων πάνω στην οποία θα γίνει η αναλυτική επεξεργασία δεδομένων (OLAP) και η οπτικοποίηση των αποτελεσμάτων σε χάρτη GIS. Ακολούθως αναλύονται οι διαδικασίες υλοποίησης της προαναφερθείσας αποθήκης δεδομένων και της αρχιτεκτονικής αυτής, ενώ οι δύο τελευταίες ενότητες παρουσιάζουν τις ενέργειες που πραγματοποιήθηκαν στο σύστημα και στα δεδομένα για να είναι δυνατή η OLAP ανάλυση και η GIS οπτικοποίηση πάνω στα αποτελέσματα του OLAP.

Η αποθήκη δεδομένων θα «γεμίσει» και από στοιχεία των αλλοδαπών ασφαλισμένων του ΙΚΑ που παρέχονται από την Καρασιώτου Βάσω.

5.1 Αρχιτεκτονική συστήματος

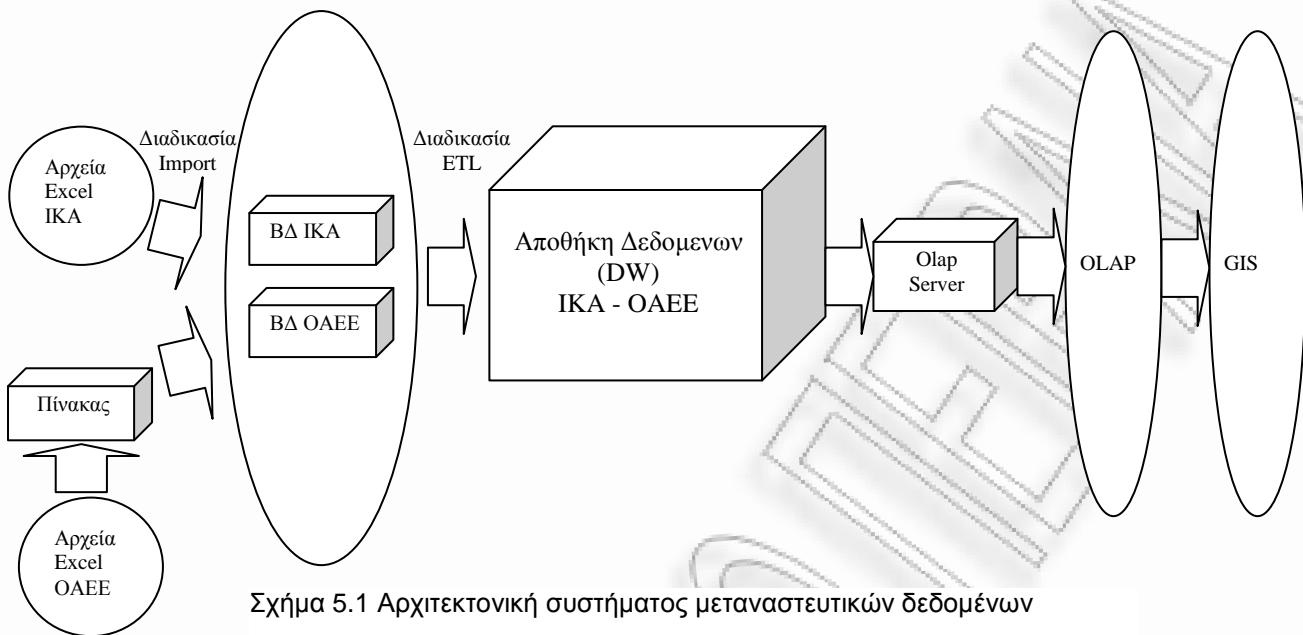
Η αρχιτεκτονική του συστήματος φαίνεται στο σχήμα 5.1. Τα αρχικά δεδομένα του ΙΚΑ είναι σε μορφή Excel και εισάγονται στη βάση δεδομένων του ΙΚΑ. Τα αντίστοιχα του ΟΑΕΕ θα είναι και αυτά σε μορφή excel, η ανάλυση των οποίων παρουσιάζεται στην επόμενη ενότητα. Αρχικά θα γίνει εισαγωγή των δεδομένων, με την ακριβή μορφή που θα είναι, σε έναν πίνακα του περιβάλλοντος Microsoft SQL Server 2005. Έτσι αν η ημερομηνία γέννησης είναι σε πεδίο χαρακτήρων στο νέο πίνακα θα εισαχθεί με την ακριβή αυτή μορφή και όχι μετατρεπόμενο σε πεδίο ημερομηνίας. Για τη διαδικασία του import από το excel στον αρχικό πίνακα θα χρησιμοποιηθεί σαν ενδιάμεσο εργαλείο η Microsoft Access 2003. Αρχικά θα φορτωθούν τα δεδομένα από το excel στην access με το ενσωματωμένο εργαλείο και ακολούθως θα γίνει εξαγωγή των δεδομένων από την access στον SQL Server μέσω ODBC. Η ίδια διαδικασία πραγματοποιείται και για άλλα δεδομένα, όπως οι ταχυδρομικοί κώδικες της Ελλάδας, που είναι και αυτά σε μορφή excel.

Ακολούθως πραγματοποιείται η υλοποίηση μιας βάσης δεδομένων για υποδοχή των δεδομένων του νέου πίνακα σε κανονικοποιημένη μορφή και πάλι σε περιβάλλον Microsoft SQL Server 2005, όπως φαίνεται στην ενότητα 5.2. Για τη μεταφορά των δεδομένων από τον προσωρινό πίνακα στην κανονικοποιημένη βάση δεδομένων χρησιμοποιείται απλή SQL και όπου χρειαστεί T-SQL, μία επέκταση της SQL σε περιβάλλον SQL Server. Τα scripts που χρησιμοποιήθηκαν παρατίθενται στο παράρτημά. Η εισαγωγή των δεδομένων στη βάση του ΙΚΑ και η υλοποίηση της βάσης δεδομένων του ΙΚΑ είναι αντικείμενο άλλης διπλωματικής.

Η αποθήκη δεδομένων που δημιουργήθηκε θα δεχθεί δεδομένα τόσο από τη βάση δεδομένων του ΙΚΑ, όσο και του ΟΑΕΕ, όπως παρουσιάζεται στην ενότητα 5.3. Συγκεκριμένα βλέπουμε την υλοποίηση της αποθήκης δεδομένων, ενώ αναλύονται κάποιες ETL διαδικασίες για τη μεταφορά των δεδομένων από τη βάση δεδομένων του ΟΑΕΕ στην αποθήκη δεδομένων. Κάποιες ETL διαδικασίες περιέχουν παράλληλη εισαγωγή και των δεδομένων του ΙΚΑ, ενώ όποια διαδικασία δεν αφορούσε τα δεδομένα του ΟΑΕΕ δεν παρουσιάζεται στη συγκεκριμένη διπλωματική. Η υλοποίηση της αποθήκης δεδομένων έχει πραγματοποιηθεί σε περιβάλλον Sql Server 2005 και η ETL είτε σε SQL είτε σε T-SQL.

Για τη δημιουργία του κύβου, ο οποίος θα διαχειριστεί τα δεδομένα της αποθήκης δεδομένων χρησιμοποιήθηκε το Microsoft Analysis Services 2005. Στην ενότητα 5.4 παρουσιάζονται οι διαστάσεις και τα μέτρα του κύβου, πάνω στον οποίο θα γίνει OLAP ανάλυση.

Τέλος για την οπτικοποίηση των OLAP αποτελεσμάτων θα χρησιμοποιηθεί το ArcGIS 9.1 και το Add-on OLAP for ARCGIS, μέσω του οποίου θα δημιουργούμε από το περιβάλλον του ArcGIS OLAP ερωτήματα, τα οποία ακολούθως θα οπτικοποιηθούν σε χάρτες.



Σχήμα 5.1 Αρχιτεκτονική συστήματος μεταναστευτικών δεδομένων

5.2 Σχεδιασμός και υλοποίηση της Βάσης Δεδομένων του ΟΑΕΕ

Για την υλοποίηση της βάσης δεδομένων του ΟΑΕΕ θα χρειαστεί να δοθούν κάποια στοιχεία από το φορέα. Επειδή τα στοιχεία αυτά είναι «Προσωπικά δεδομένα» δεν είναι άμεσα διαθέσιμα για αυτή τη διπλωματική, αλλά ο συγγραφέας γνωρίζει τις αδυναμίες και τους περιορισμούς των στοιχείων αυτών, λόγω ενεργούς συμμετοχής του σε ομάδα εργασίας για την εκκαθάριση μητρώου ασφαλισμένων του ΟΑΕΕ. Το γεγονός αυτό δεν καθιστά αναγκαία την παροχή των στοιχείων αυτών για τη δημιουργία της βάσης δεδομένων, αλλά ούτε και για τη δημιουργία των διαδικασιών για την εισαγωγή των στοιχείων σε αυτήν, οι οποίες φαίνονται στο παράρτημα.

Τα στοιχεία αυτά για τον κάθε ασφαλισμένο θα είναι τα εξής:

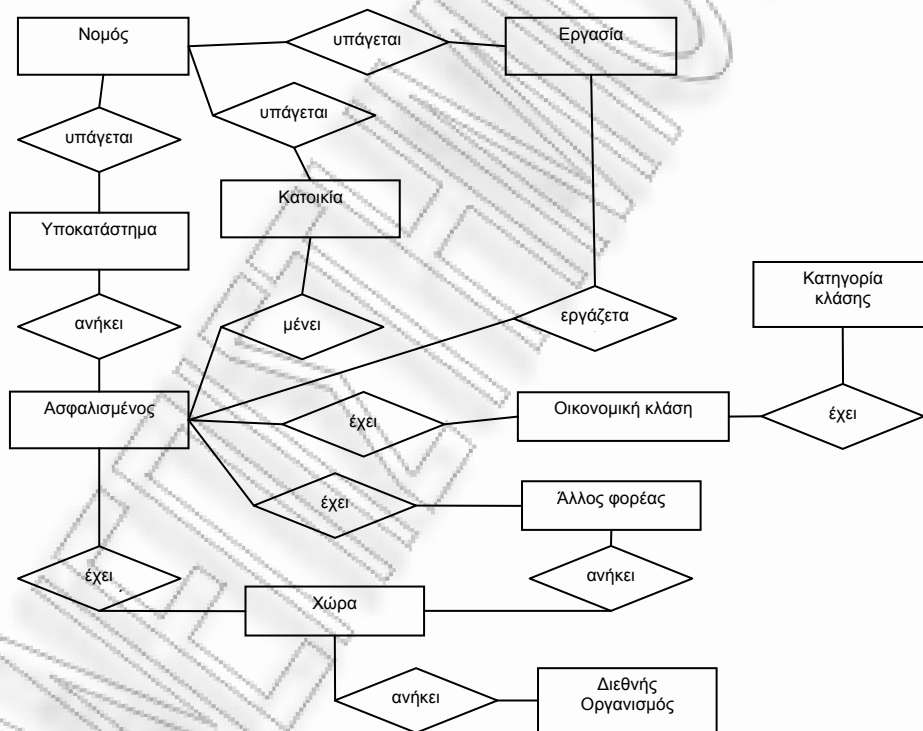
- Έτος γέννησης
- Μήνας γέννησης
- Μέρα γέννησης
- Φύλο
- Οικογενειακή κατάσταση
- Επάγγελμα
- Ασφαλιστική κατηγορία
- Ημ. έναρξης ασφάλισης
- Ημ. διακοπής
- Ημ. επανεγγραφής
- Χώρα γέννησης
- ΤΚ επιχείρησης
- ΤΚ κατοικίας
- Προστατευόμενα μέλη
- Άλλος φορέας ασφάλισης

Από τα ανωτέρω θα δημιουργηθεί μία βάση δεδομένων για τον ΟΑΕΕ με τις παρακάτω απαιτήσεις και περιορισμούς.

5.2.1 Απαιτήσεις/ Περιορισμοί – ER

1. Κάθε ασφαλισμένος έχει μία χώρα γέννησης
2. Κάθε χώρα ανήκει σε ένα ή περισσότερους οργανισμούς (ΕΕ , NATO ,κλπ.)
3. Κάθε ασφαλισμένος μένει σε μία διεύθυνση
4. Κάθε διεύθυνση ανήκει σε ένα νομό
5. Κάθε ασφαλισμένος έχει μία έδρα εργασίας
6. Κάθε έδρα εργασίας ανήκει σε ένα υποκατάστημα του οργανισμού
7. Κάθε υποκατάστημα ανήκει σε ένα νομό
8. Κάθε ασφαλισμένος ανήκει σε ένα ή περισσότερους φορείς (εντός ή εκτός Ελλάδος)
9. Κάθε φορέας ανήκει σε μία χώρα
10. Κάθε ασφαλισμένος έχει μία οικονομική κλάση ανά μήνα ενός έτους
11. Κάθε οικονομική κλάση ανήκει σε μία κατηγορία
12. Κάθε οικονομική κλάση έχει ένα ποσό ανά έτος

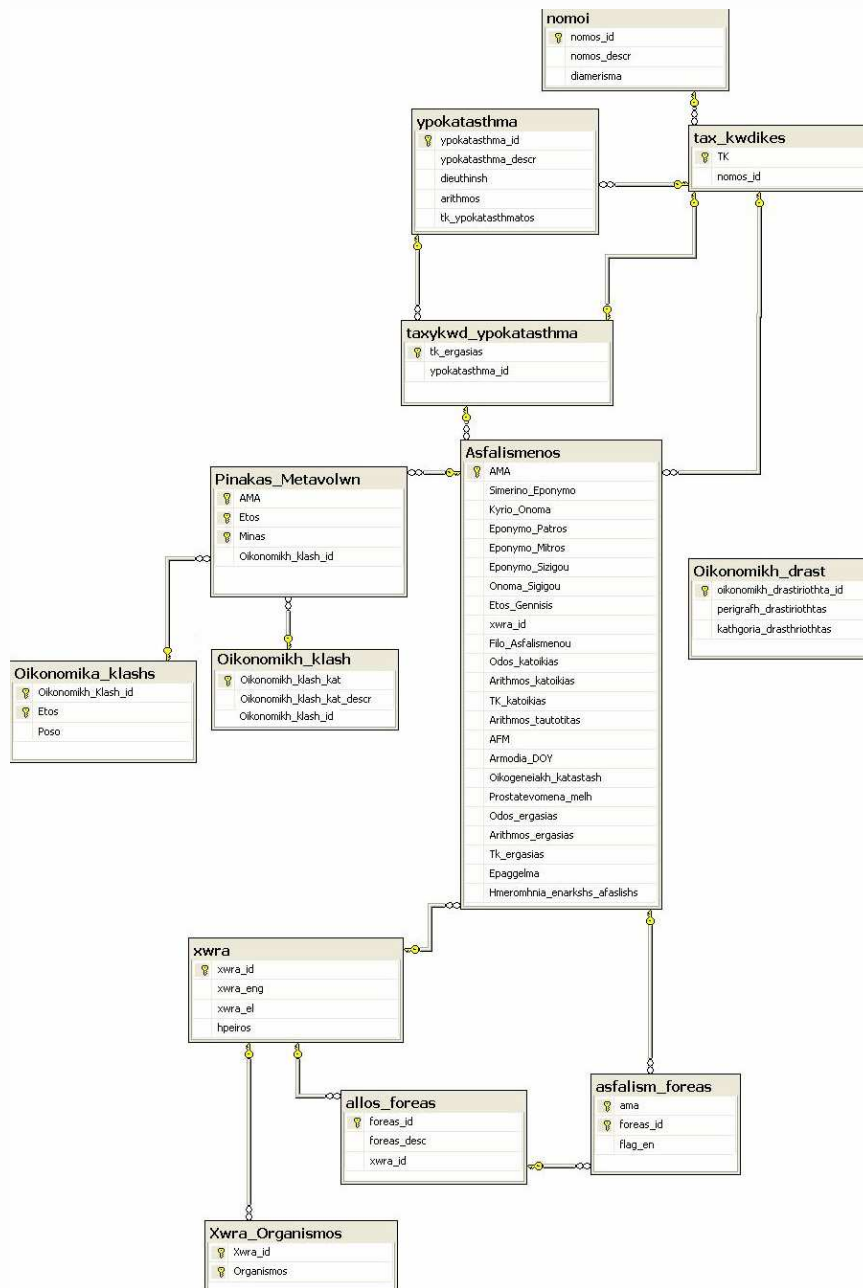
Από τις απαιτήσεις αυτές συναρτησει των δεδομένων μπορούμε να εξάγουμε το ER της βάσης δεδομένων, το οποίο φαίνεται στο σχήμα 5.2.



Σχήμα 5.2 ER της Βάσης Δεδομένων

5.2.2 Σχεσιακό διάγραμμα – Ανάλυση βάσης δεδομένων

Από το λογικό διάγραμμα που φαίνεται στο ανωτέρω σχήμα μπορούμε να εξάγουμε το σχεσιακό διάγραμμα της βάσης δεδομένων, το οποίο παρουσιάζεται στο σχήμα 5.3.



5.3 Σχισιακό διάγραμμα της Βάσης Δεδομένων

Ο πίνακας **TAXYKWD_YPOKATASTHMA** περιέχει τους ταχυδρομικούς κώδικες στους οποίους δραστηριοποιείται ένας ασφαλισμένος και ποιους εξυπηρετεί το κάθε υποκατάστημα. Συνδέεται με τον ταχυδρομικό κωδικό εργασίας του πίνακα **ASFALISMENOS**. Παράλληλα συνδέεται με τους πίνακες **YPOKATASTHMA** και **TAX_KWDIKES**. Ο συγκεκριμένο πίνακας βρέθηκε από τον παλιό ιστότοπο του ΟΑΕΕ.

Ο πίνακας **YPOKATASTHMA** περιέχει τα υποκαταστήματα του ΟΑΕΕ και χαρακτηριστικά αυτών, όπως η περιγραφή του και η διεύθυνση στην οποία βρίσκεται. Ο ταχυδρομικός κώδικας που περιέχεται στον πίνακα είναι η σύνδεση με τον πίνακα **TAX_KWDIKES**. Και αυτός ο πίνακας βρέθηκε από τον παλιό ιστότοπο του ΟΑΕΕ.

Από τον ιστότοπο <http://geodata.gov.gr/geodata/> βρέθηκαν σε αρχείο μορφής excel οι

ταχυδρομικοί κώδικες της Ελλάδας και το λεκτικό του νομού που ανήκουν. Από το αρχείο αυτό δημιουργήθηκαν οι δύο παρακάτω αναλυόμενοι πίνακες, TAX_KWDIKES και NOMOI.

Ο πίνακας **TAX_KWDIKES** περιέχει τους ταχυδρομικούς κώδικες της Ελλάδος και την κωδικοποίηση του νομού στον οποίο ανήκουν.

Ανάλογα ο πίνακας **NOMOI** περιέχει τους νομούς της Ελλάδας. Η κωδικοποίηση των νομών είναι ένας τυχαίος αριθμός ώστε να συσχετιστούν οι πίνακες NOMOI και TAX_KWDIKES.

Ο πίνακας **XWRA** περιέχει το διψήφιο iso μια χώρας την περιγραφή της σε αγγλικά και ελληνικά και την ήπειρο στην οποία ανήκει. Ο συγκεκριμένος πίνακας βρέθηκε από την wikipedia.

Καθώς μία χώρα μπορεί να ανήκει σε πολλούς οργανισμούς (ΕΕ, Ευρωζώνη, NATO, ΟΗΕ) υπήρξε η ανάγκη του πίνακα **XWRA_ORGANISMOS** όπου υπάρχουν οι χώρες με το διψήφιο iso τους και ο οργανισμός (ένας ή πολλοί) στον οποίο ανήκουν. Συγκεκριμένα εδώ έχουν καταχωρηθεί οι χώρες που ανήκουν στην ΕΕ.

Ο πίνακας **ΟΙΚΟΝΟΜΙΚΗ_DRAST** παρόλο που θα ήταν εξαιρετικά χρήσιμος δεν κατέστη δυνατόν να υλοποιηθεί, λόγω της έλλειψης κωδικοποίησης επαγγελμάτων από τον ΟΑΕΕ.

Ο πίνακας **ΟΙΚΟΝΟΜΙΚΑ_KLASHS** περιέχει το ποσό που πληρώνει ο κάθε ασφαλισμένος του ΟΑΕΕ ανάλογα την κλάση την οποία ανήκει και ποιο έτος αφορά. Ο συγκεκριμένος πίνακας βρέθηκε από τον ιστότοπο του ΟΑΕΕ.

Ο πίνακας **ΟΙΚΟΝΟΜΙΚΗ_KLASH** περιέχει την κατηγορία που ανήκει η κάθε κλάση, αν δηλαδή αυτή αναφέρεται σε νέο ή σε παλαιό ασφαλισμένο. Ο συγκεκριμένος πίνακας δε χρειάστηκε να έχει περιεχόμενα, καθώς ο ΟΑΕΕ τελικά έχει διαφορετική κωδικοποίηση αν ο ασφαλισμένος είναι νέος ή παλιός, αλλά υπάρχει η πρόβλεψη αν αυτό διαφοροποιηθεί.

Ο πίνακας **ALLOS_FOREAS** περιέχει άλλους φορείς ασφάλισης πέραν του ΟΑΕΕ, είτε στην Ελλάδα, είτε στο εξωτερικό. Ο πίνακας περιέχει την περιγραφή του φορέα και το διψήφιο iso της χώρας στην οποία ανήκει ο φορέας, οπότε και συνδέεται με τον πίνακα XWRA ο οποίος περιέχει την περιγραφή της χώρας.

Ο πίνακας **ASFALISM_FOREAS** περιέχει τον κωδικό του ασφαλισμένου και τον κωδικό του άλλου φορέα στον οποίο ανήκει, καθώς και μία ένδειξη αν αυτός είναι ενεργός ή όχι.

Ο πίνακας **PINAKAS_METAVOLWN** θα υλοποιηθεί συναρτήσει των παρεχόμενων στοιχείων και ανάλογα την ημερομηνία ασφάλισης, διακοπής και επανεγγραφής. Εδώ περιέχεται ο αριθμός μητρώου του ασφαλισμένου, το έτος και ο μήνας που αφορά την εγγραφή και η οικονομική κλάση στην οποία υπάγεται ο ασφαλισμένος το συγκεκριμένο μήνα. Ο κώδικας που χρησιμοποιήθηκε για την εισαγωγή των δεδομένων του εν λόγω πίνακα είναι T-SQL και παρέχεται στο παράρτημα.

Ο πίνακας **ASFALISMENOS** θα περιέχει τα στοιχεία του ασφαλισμένου, δημογραφικά και επαγγελματικά. Το πεδίο tk_ergasias θα μπορούσε να βρίσκεται στον πίνακα PINAKAS_METAVOLWN καθώς θα ήταν ενδιαφέρον να υπάρχουν στατιστικά δεδομένα αλλαγής τόπου εργασίας. Αυτό όμως δεν είναι δυνατόν λόγω του ότι δεν διατηρούνται οι μεταβολές του εν λόγω πεδίου αλλά μόνο η τελευταία τιμή.

5.3 Αποθήκη Δεδομένων

Ο απώτερος στόχος στην υλοποίηση της συγκεκριμένης αποθήκης δεδομένων είναι με τα κατάλληλα εργαλεία να γίνει εφικτή η λήψη στρατηγικών αποφάσεων. Χαρακτηριστικό παράδειγμα θα ήταν η στοχευμένη πάταξη της εισφοροδιαφυγής.

Για τη δημιουργία της «Αποθήκης Δεδομένων» χρησιμοποιήθηκε το σχήμα χιονονιφάδας «Snowflake», μία βελτίωση του αστεροειδούς σχήματος, όπου η ιεραρχία των διαστάσεων αναπαρίσταται κανονικοποιώντας τους πίνακες των διαστάσεων. Η λογική δημιουργίας αυτής καθορίστηκε τόσο από τον παραπάνω στόχο όσο και από τα πρωτογενή δεδομένα, τα οποία καθόρισαν του πίνακες διαστάσεων, τους πίνακες γεγονότων αλλά και τα μετρήσιμα μεγέθη.

Ακολούθως εισήχθησαν τα δεδομένα από τις βάσεις δεδομένων του ΙΚΑ – ΕΤΑΜ και του ΟΑΕΕ/ ΤΕΒΕ, αφού αρχικά ενοποιηθήκαν πιθανές διαφοροποιήσεις κωδικών σε χώρες και σε νομούς. Επίσης ορίστηκε και ο πίνακας DIM_TIME με κωδικοποίηση των ημερομηνιών ανά εξάμηνο και έτος, οπότε τα δεδομένα που εισήχθησαν στον πίνακα

FACT_PERIODIKH_KINHSH είναι αθροίσματα των μηνιαίων ποσών που υπήρχαν στη βάση δεδομένων του ΟΑΕΕ/ΤΕΒΕ ανά εξάμηνο. Ακολουθούν οι διαστάσεις και τα μέτρα της «Αποθήκης Δεδομένων», ενώ στην επόμενη ενότητα παρουσιάζεται το διάγραμμα της αποθήκης δεδομένων και αναλύονται οι πίνακες που την αποτελούν..

Διαστάσεις της αποθήκης δεδομένων

Χρόνος

Έτος → Εξάμηνο

Οικονομική Δραστηριότητα

Περιοχή Εργασίας

Γεωγραφικό διαμέρισμα → Νομός

Περιοχή Κατοικίας

Γεωγραφικό διαμέρισμα → Νομός

Οικογενειακή Κατάσταση

Καταγωγή

Ηπειρος χώρας καταγωγής → Οργανισμός χώρας καταγωγής → Χώρα καταγωγής

Υποκατάστημα Φορέα

Φορέας → Υποκατάστημα Φορέα

Φύλο

Έτος Απογραφής

Έτος Γέννησης

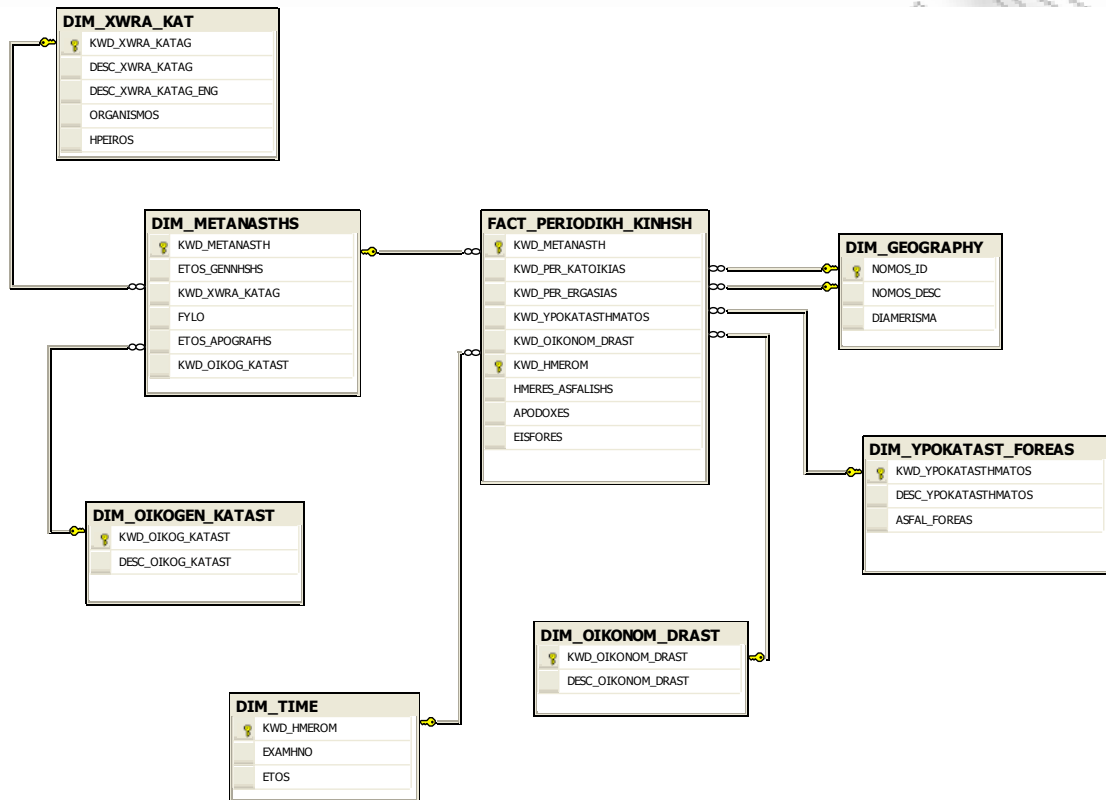
Μέτρα

Πλήθος εισφορών

Πλήθος αποδοχών

Αριθμός μεταναστών

5.3.1 Σχεδίαση της ΑΔ



Σχήμα 5.4 Σχεσιακό διάγραμμα αποθήκης δεδομένων

Ο πίνακας **FACT_PERIODIKH_KINHSH** περιέχει τις μεταβολές του ασφαλισμένου ανά εξάμηνο, όπως εισήχθησαν από τις βάσεις δεδομένων.

Περιέχονται μέτρα, όπως οι ημέρες ασφάλισης (**HMERES_ASFALISHS**), οι αποδοχές (**APODOXES**) και οι εισφορές (**EISFORES**). Οι ημέρες ασφάλισης και οι αποδοχές υλοποιήθηκαν για του ασφαλισμένους του ΙΚΑ, παρόλο που δεν υπήρχαν τα συγκεκριμένα δεδομένα σαν πρόβλεψη για μελλοντική παροχή αυτών. Αντίθετα το συνολικό ποσό των εισφορών που πλήρωσε ο ασφαλισμένος του ΟΑΕΕ το συγκεκριμένο εξάμηνο παρήχθη συναρτήσει των οικονομικών κλάσεων που άνηκε. Περιέχει τα κλειδιά των πινάκων διαστάσεων που συνδέονται με αυτόν όπως (**KWD_METANASTH**), (**KWD_PER_KATOIKIAS**), (**KWD_PER_ERGASIAS**), (**KWD_YPOKATASTHMATOS**). Τα δεδομένα του ΟΑΕΕ προήλθαν από ερώτημα στους πίνακες (**PINAKAS_METAVOLWN**), (**ASFALISMENOS**), (**OIKONOMIKA_KLASHS**) για την παροχή του (**EISFORES**), αλλά κ των πινάκων (**NOMOI**), (**YPOKATASTHMA**), (**TAX_KWDIKES**), (**TAXYKWD_YPOKATASTHMA**) από τους οποίου προήλθαν οι συσχετίσεις με τους πίνακες διαστάσεων. Η διάσταση (**KWD_OIKONOM_DRAST**) δεν περιέχει δεδομένα για τους ασφαλισμένους του ΟΑΕΕ, όπως αναφέρεται κα στην ανάλυση της βάσης δεδομένων. Καθώς δεν υπήρχαν στοιχεία για τα υποκαταστήματα του ΙΚΑ δημιουργήθηκε κωδικοποίηση για τον (**KWD_YPOKATASTHMATOS**) με στοιχείο '0' για τους ασφαλισμένους του ΙΚΑ, το οποίο εισήχθη και στον πίνακα (**DIM_YPOKATAST_FOREAS**) που αναλύεται παρακάτω. Στους (**KWD_PER_KATOIKIAS**) και στον (**KWD_PER_ERGASIAS**) υπάρχει η κωδικοποίηση του ΙΚΑ. Για να αντιστοιχηθεί με την κωδικοποίηση του ΟΑΕΕ χρησιμοποιήθηκε σαν κλειδί το λεκτικό κάθε νομού. Όπου υπήρξε αναντιστοιχία υπήρξε μεταβολή της τιμής του λεκτικού στη βάση δεδομένων του ΟΑΕΕ, ώστε να ταυτίζεται με αυτή του ΙΚΑ.

Ο πίνακας **DIM_GEOGRAPHY** περιέχει την κωδικοποίηση των νομών του ΙΚΑ ενώ το πεδίο (**DIAMERISMA**) εισήχθη από τη βάση δεδομένων του ΟΑΕΕ για τον ανάλογο νομό. Επειδή στη βάση δεδομένων του ΟΑΕΕ υπάρχει η κωδικοποίηση 'ΑΤΤΙΚΗ' για το νομό Αττικής,

ενώ στου ΙΚΑ υπάρχει η διοικητική διαίρεση του νομού Αττικής στις νομαρχίες Ανατολικής και Δυτικής Αττικής, Πειραιά και Αθηνών έγινε ενοποίηση της διοικητικής διαίρεσης σε μία εγγραφή στην 'ΑΤΤΙΚΗ' στους ασφαλισμένους του ΙΚΑ.

Ο πίνακας **DIM_XWRA_KAT** περιέχει την κωδικοποίηση των χωρών σύμφωνα με το iso τους (KWD_XWRA_KATAG), το λεκτικό τους στα ελληνικά (DESC_XWRA_KATAG), το λεκτικό τους στα Αγγλικά (DESC_XWRA_KATAG_ENG), καθώς και την ήπειρο που ανήκει η χώρα (HPEIROS). Στο πεδίο (ORGANISMOS) περιέχεται η ένδειξη ΕΕ ώστε να ξεχωρίζουν οι χώρες που βρίσκονται στην ευρωπαϊκή ένωση. Δεν υπήρξαν προβλήματα στη σύνδεση ανάμεσα στις δύο βάσεις δεδομένων, καθώς ο κωδικός της κάθε χώρας ήταν το διψήφιο iso αυτής και στου δύο φορείς.

Ο πίνακας **DIM_METANASTHS** περιέχει τους ασφαλισμένους του ΙΚΑ και το ΟΑΕΕ με την κωδικοποίηση που τους έχει δοθεί στις βάσεις δεδομένων του. Περιέχει τα στοιχεία που δεν αλλάζουν για ένα φυσικό πρόσωπο όπως το έτος γέννησης (ETOS_GENNHSHS), η χώρα καταγωγής (KWD_XWRA_KATAG), το φύλο (FYLO) και το έτος απογραφής (ETOS_APOGRAFHS), ενώ περιέχει και στα στοιχεία την τελευταία οικογενειακή δραστηριότητα (KWD_OIKOG_KATAST) παρόλο που αυτή μπορεί να αλλάζει, λόγω των αρχικών δεδομένων και της μη διαφοροποίησης αυτής στους ασφαλισμένους του ΟΑΕΕ, ενώ στους ασφαλισμένους του ΙΚΑ δεν υπάρχουν διαθέσιμα στοιχεία οικογενειακής κατάστασης.

Ο πίνακας **DIM_OIKOG_KATAST** περιέχει την κωδικοποίηση της οικογενειακής κατάστασης κάθε ασφαλισμένου. Για το ΙΚΑ δεν υπάρχουν στοιχεία.

Ο πίνακας **DIM_TIME** περιέχει τη διάσταση του χρόνου, που είναι απαραίτητο στοιχείο κάθε αποθήκης δεδομένων. Περιέχει κωδικοποίηση ημερομηνίας για τα έτη από το 2004 έως το 2009 και για μονό ή ζυγό εξάμηνο.

Ο πίνακας **DIM_OIKONOM_DRAST** περιέχει την κωδικοποίηση της οικονομικής δραστηριότητας των ασφαλισμένων του ΙΚΑ. Για τον ΟΑΕΕ δεν υπάρχει πληροφορία.

Ο πίνακας **DIM_YPOKATAST_FOREAS** περιέχει το τμήμα στο οποίο υπάγεται ο κάθε ασφαλισμένος του ΟΑΕΕ. Για το ΙΚΑ δεν υπάρχει αντίστοιχη πληροφορία, οπότε στον πίνακα υπάρχει ένα record που αντιστοιχεί στην κωδικοποίηση ενός τμήματος για όλους τους ασφαλισμένους του ΙΚΑ.

5.4 Δημιουργία κύβου για OLAP ανάλυση

Στην παρούσα ενότητα παρουσιάζεται η προετοιμασία του κατάλληλου περιβάλλοντος ώστε να υλοποιηθεί η αναλυτική επεξεργασία δεδομένων. Πραγματοποιείται η δημιουργία του κύβου δεδομένων, ενώ προετοιμάζονται τα μέτρα και οι διαστάσεις αυτού. Η παρουσίαση γίνεται στο Business Intelligence Management Studio του Microsoft Analysis Services 2005.

Μετά τη δημιουργία της αποθήκης δεδομένων και για την παρασκευή του κύβου αναγνωρίστηκε η ανάγκη δημιουργίας δύο επιπλέον διαστάσεων που θα αντικαθιστούσαν τις διαστάσεις «Έτος Απογραφής» και «Έτος γέννησης» με μία περαιτέρω ανάλυση αυτών, η οποία είναι η δεκαετία στην οποία ανήκει το κάθε έτος. Για την επίλυση αυτού δημιουργήθηκαν 2 views (όψεις) στην αποθήκη δεδομένων. Η όψη (VW_DEKAETIA) που περιέχει το κάθε έτος που υπάρχει στον πίνακα (DIM_METANASTHS) και τη δεκαετία στην οποία αντιστοιχεί (π.χ. 1932, 1930-1939), και η όψη (METANASTHS_ME_DEKAETIA) που περιέχει τα στοιχεία που περιέχονται στον πίνακα (DIM_METANASTHS), αλλά το (ETOS_GENNHSHS) και το (ETOS_APOGRAFHS) είναι κωδικοποιημένα και αντιστοιχούν στην όψη (VW_DEKAETIA). Έτσι αντί για τις διαστάσεις Έτος Απογραφής και Έτος Γέννησης έχουμε τις:

Απογραφή ασφαλισμένου

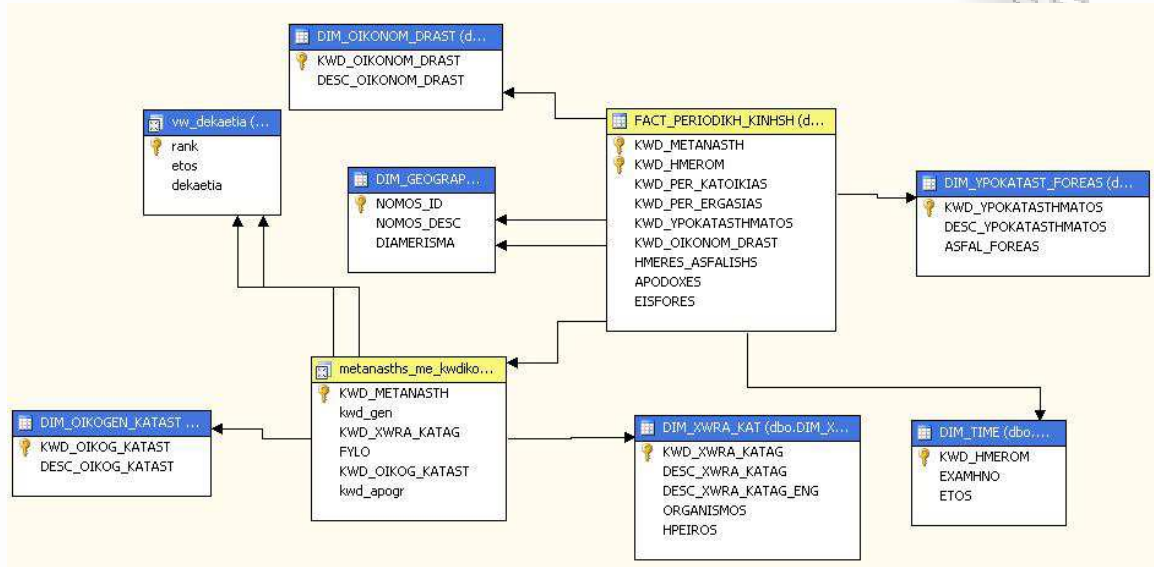
Δεκαετία απογραφής → Έτος απογραφής

Γέννηση ασφαλισμένου

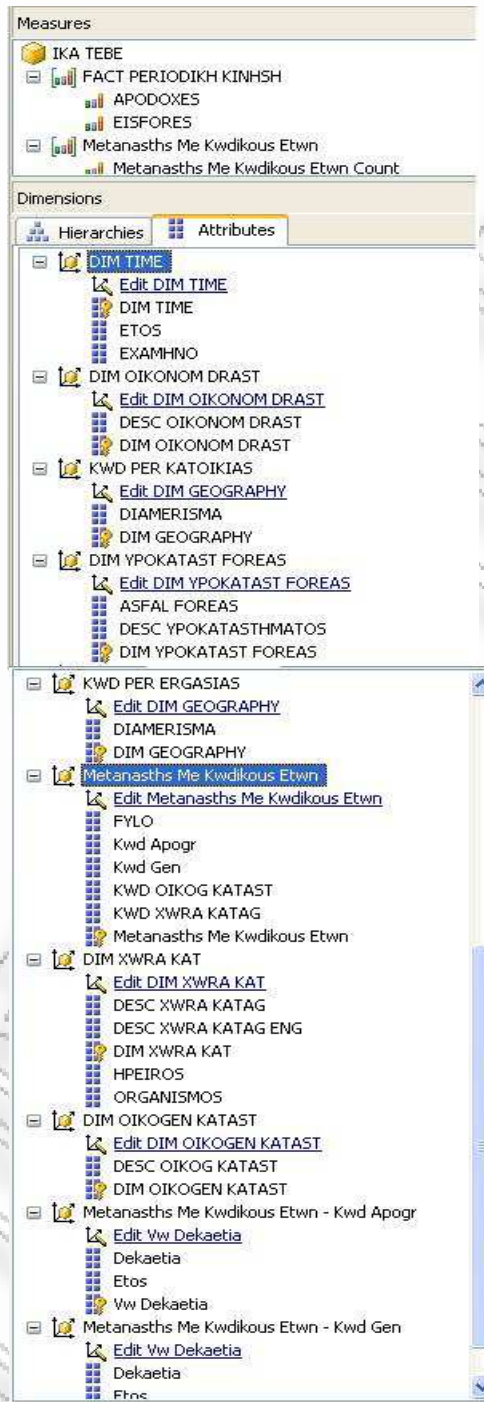
Δεκαετία γέννησης → Έτος γέννησης

Οπότε σύμφωνα με τα ανωτέρω το μοντέλο του κύβου που δημιουργήθηκε φαίνεται στο

σχήμα 5.5. και οι διαστάσεις και τα μέτρα στο σχήμα 5.6



Σχήμα 5.5 Κύβος δεδομένων



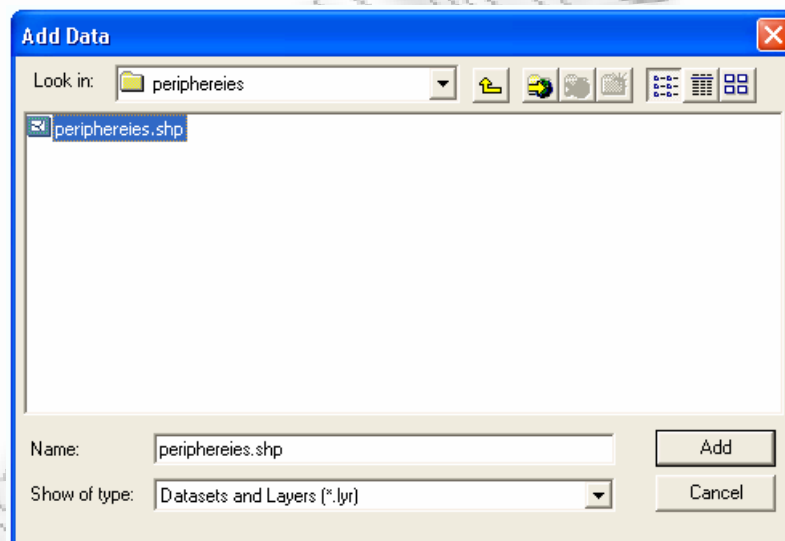
Σχήμα 5.6 Μέτρα και διαστάσεις κύβου δεδομένων

5.5 Δημιουργία χαρτών για οπτικοποίηση διαφόρων ερωτημάτων σε GIS

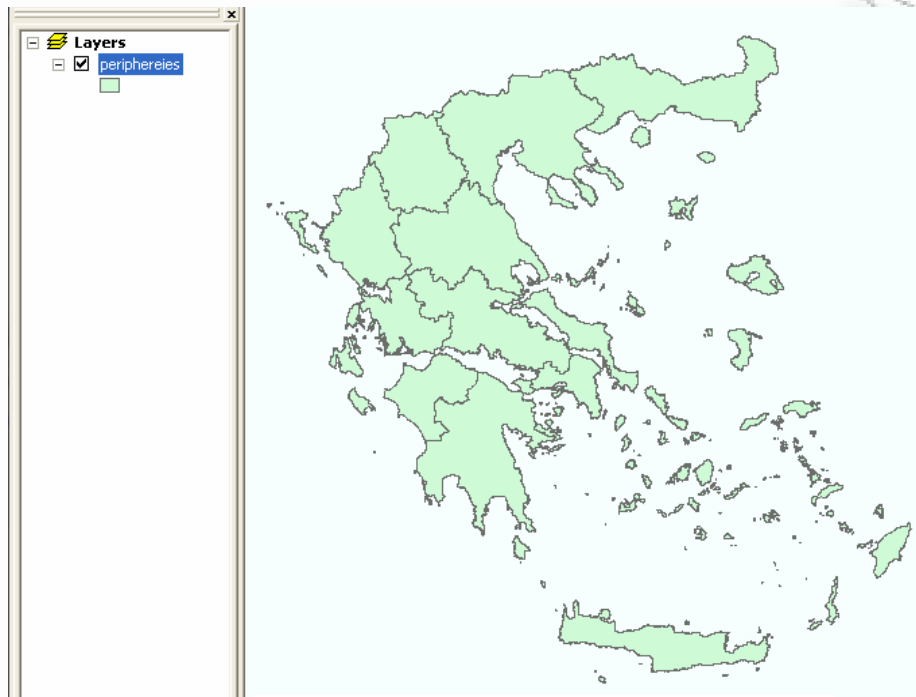
Στόχος της συγκεκριμένης ενότητας είναι η προετοιμασία του GIS περιβάλλοντος (ARCGIS), ώστε να επιτευχθεί η οπτικοποίηση των αποτελεσμάτων των OLAP ερωτημάτων σε GIS χάρτες. Παράλληλα λόγω της ανάγκης σύνδεσης των χαρτών με τα αποτελέσματα του OLAP ίσως χρειαστεί η μετατροπή κάποιων δεδομένων τόσο των χαρτών, όσο και της αποθήκης δεδομένων.

Στην ηλεκτρονική διεύθυνση <http://geodata.gov.gr/geodata/>, βρίσκονται ελεύθερα ανοιχτά γεωγραφικά δεδομένα της ελληνικής δημόσιας διοίκησης. Το geodata.gov.gr αποτελεί την πρώτη προσπάθεια για τη δωρεάν διάθεση γεωχωρικών δεδομένων της ευρύτερης Δημόσιας Διοίκησης προς όλους τους πολίτες της χώρας. Από το συγκεκριμένο ιστότοπο βρέθηκαν σε shapefile οι νομοί της Ελλάδας και οι Περιφέρειες της Ελλάδας. Δεν κατέστη δυνατό να βρεθούν σε shapefile τα διαμερίσματα της χώρας, οπότε είναι αναγκαία η μετατροπή των περιεχομένων του πεδίου DIAMERISMA στον πίνακα DIM_GEOGRAPHY της Αποθήκης Δεδομένων. Οπότε κάθε νομός πλέον θα αντιστοιχεί σε κάποια περιφέρεια της χώρας και όχι σε κάποιο γεωγραφικό διαμέρισμα. Η ονομασία του πεδίου θα παραμείνει η ίδια.

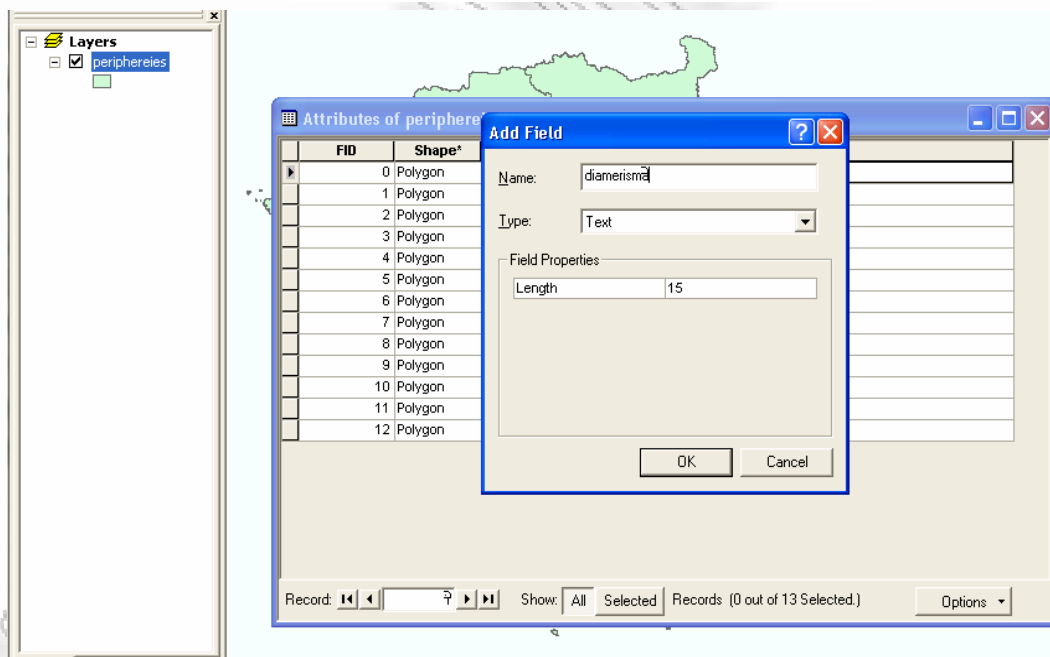
Παράλληλα θα χρειαστεί να διαφοροποιηθούν τα δεδομένα του shapefile ώστε να ταυτίζονται με τα δεδομένα του κύβου για να γίνει η συσχέτιση ανάμεσα σε shapefile και κύβο. Στα επόμενα σχήματα βλέπουμε την εισαγωγή του shapefile των περιφερειών στο χάρτη και την τροποποίηση των δεδομένων αυτού.



Σχήμα 5.7 Εισαγωγή shapefile στον χάρτη



Σχήμα 5.8 Οπτικοποίηση shapefile στον χάρτη



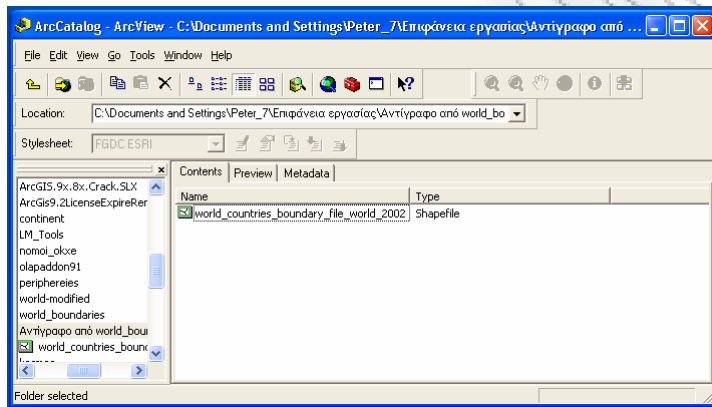
Σχήμα 5.9 Εισαγωγή πεδίου που θα χρησιμοποιηθεί σαν κλειδί με το αποτέλεσμα του OLAP

Ακολούθως γίνεται καταχώρηση των περιφερειών της Ελλάδας στο πεδίο που δημιουργήθηκε, οπότε θα χρησιμοποιηθεί σαν συσχέτιση με το OLAP αποτέλεσμα. Τελειώνοντας αποθηκεύουμε το χάρτη ως "Perifereies.mxd", ο οποίος περιέχει το συγκεκριμένο shapefile.

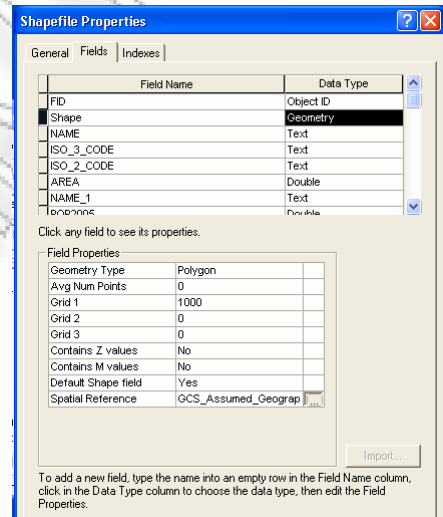
Αντίστοιχες ενέργειες γίνονται και για το shapefile με τους νομούς της Ελλάδας, καθώς έπρεπε να εισαχθεί στο συγκεκριμένο shapefile η κωδικοποίηση των νομών που έχει επιλεγεί στην αποθήκη δεδομένων στον κάθε νομό που υπάρχει στο shapefile. Και εδώ αποθηκεύουμε το χάρτη ως "Nomi.mxd", ο οποίος περιέχει το shapefile με τους νομούς της Ελλάδος και θα

γίνει αργότερα σε αυτόν και η συσχέτιση με τα αποτελέσματα του OLAP.

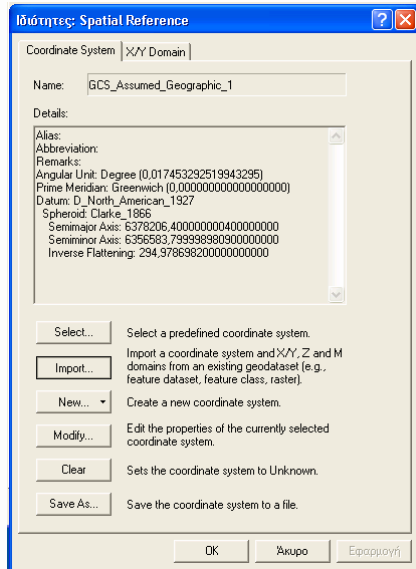
Το shapefile για τις χώρες όλου του κόσμου βρέθηκε στον ιστότοπο <http://geocommons.com>. Το συγκεκριμένο shapefile περιέχει την iso κωδικοποίηση κάθε χώρας, οπότε η σύνδεση με τα αποτελέσματα από το OLAP μπορεί να γίνει αυτόματα χωρίς περαιτέρω επεξεργασία, ειδάλλως θα ήταν πολύ δύσκολο να γίνει συσχέτιση λόγω του πλήθους των χωρών που υπάρχουν. Το συγκεκριμένο shapefile έχει διαφορετικό σύστημα συντεταγμένων οπότε θα πρέπει να γίνει αλλαγή αυτού από D_North_American_1927 σε ΕΓΣΑ 87. Για τη μετατροπή θα χρησιμοποιήσουμε το ArcCatalog, ενσωματωμένο πρόγραμμα στο ArcGIS Desktop. Αφού ανοίξουμε το εν λόγω πρόγραμμα, βρίσκουμε το shapefile που θέλουμε να μετατρέψουμε. Με δεξί κλικ οδηγούμαστε στις ιδιότητές του, όπως φαίνεται στα σχήματα 5.10 και 5.11. Στην καρτέλα field στο πεδίο shape βρίσκουμε τις ιδιότητες του Spatial Reference, τις οποίες βλέπουμε στο Σχ. 5.12. Ακολούθως πατάμε το Import και ψάχνουμε να βρούμε ένα shapefile με το επιθυμητό σύστημα συντεταγμένων, π.χ. το shapefile των νομών της Ελλάδος. Στο σχ. 5.13 βλέπουμε ότι το shapefile έχει δεχθεί τις νέες συντεταγμένες. Ακολούθως το shapefile φορτώνεται στο Arcmap και αποθηκεύεται σαν χάρτης με ονομασία "Xwres.mxd".



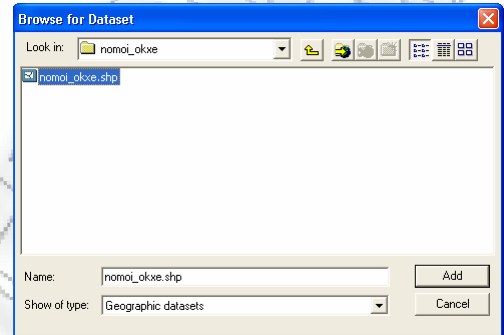
Σχ. 5.10 Εύρεση του shapefile στον arcCatalog



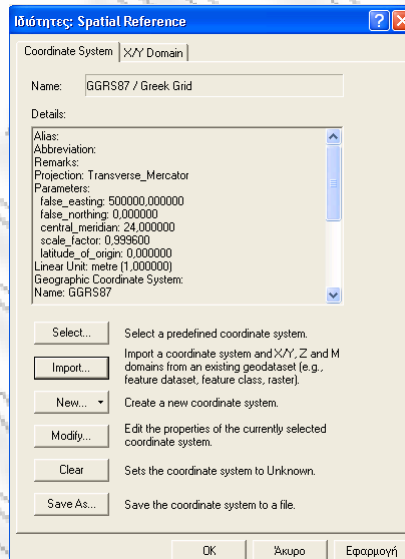
Σχ. 5.11 Ιδιότητες Shapefile



Σχ. 5.12 Ιδιότητες Spatial Reference



Σχ. 5.13 Shapefile με επιθυμητό σύστημα συντεταγμένων



Σχ. 5.14 Εισαγωγή νέου συστήματος συντεταγμένων

Εν τέλει βρέθηκε στον ιστότοπο <http://pubs.usgs.gov> το shapefile continents, το οποίο περιέχει τις ηπείρους της γης. Λόγω της έλλειψης κωδικοποίησης, αλλά και του μικρού πλήθους ηπείρων επιλέχθηκε να γίνει αλλαγή των δεδομένων του shapefile μέσα στο χάρτη, αλλάζοντας τις ονομασίες των ηπείρων στις αντίστοιχες ονομασίες που υπάρχουν στην αποθήκη δεδομένων. Η ονομασία που αποθηκεύεται αυτός ο χάρτης είναι "Continents.mxd".

6. Αναλυτική επεξεργασία μεταναστευτικών δεδομένων και οπτικοποίηση γεωγραφικής πληροφορίας

Σκοπός του παρόντος κεφαλαίου είναι η παροχή αποτελεσμάτων, μέσω OLAP και GIS περιβάλλοντος, από το σύστημα που περιγράφεται στο κεφάλαιο 5. Τα παρακάτω αποτελέσματα είναι προϊόντα του συστήματος του προηγούμενου κεφαλαίου, σύστημα το οποίο χρησιμοποιήθηκε για να δοκιμαστεί η συγκεκριμένη διαδικασία από τον ΟΑΕΕ, ώστε να φανεί κατά πόσο θα ήταν χρήσιμο για τον Οργανισμό να υλοποιηθεί το συγκεκριμένο σύστημα για το σύνολο των ασφαλισμένων του. Αυτό θα προϋπόθετε και την προμήθεια των ανάλογων πακέτων λογισμικού, είτε αυτών που χρησιμοποιούνται σε αυτή την εργασία, είτε ανταγωνιστικών αυτών. Τα αποτελέσματα θα παραδοθούν στη Διεύθυνση Στατιστικής του ΟΑΕΕ για να κρίνει τα ανωτέρω. Τα αποτελέσματα δεν παρουσιάζουν προσωπικά δεδομένα, καθώς το μόνο που παρουσιάζεται είναι στατιστικά αποτελέσματα.

Σε όποιο αποτέλεσμα υπάρχει γεωγραφική πληροφορία (νομός, χώρα κλπ.) υπάρχει κ ο αντίστοιχος GIS χάρτης.

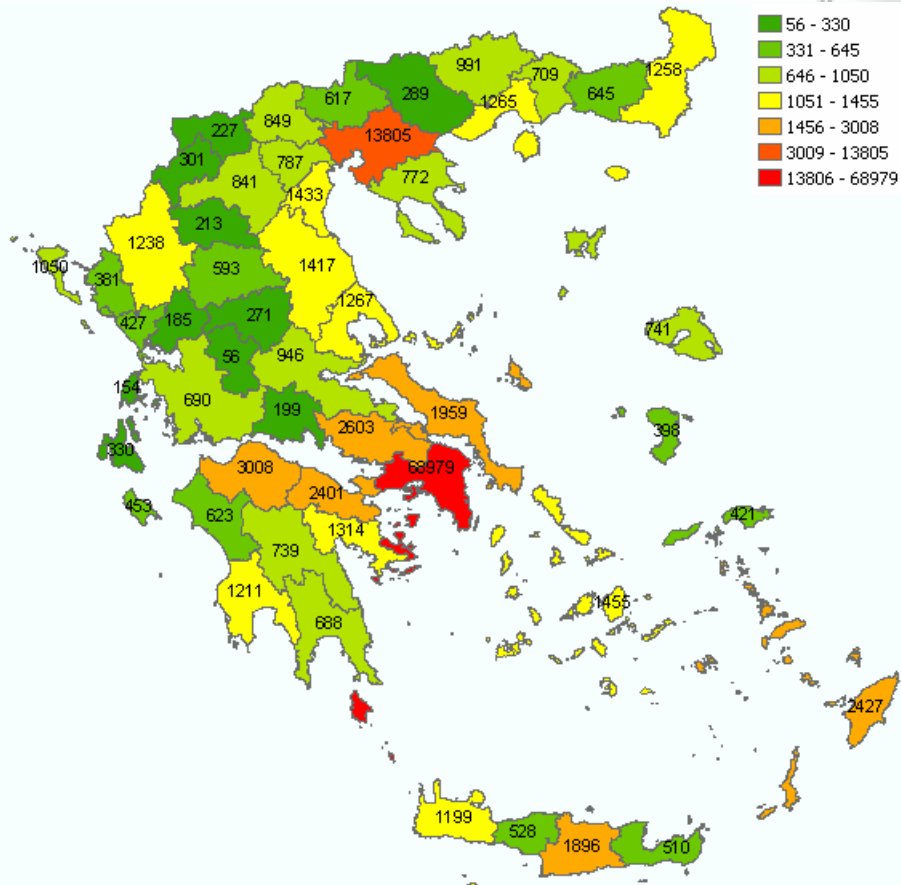
6.1 Εύρεση πλήθους μεταναστών ανά Έτος και ανά Νομό διαμονής

Στο συγκεκριμένο παράδειγμα μας ενδιαφέρει να μελετήσουμε το πλήθος των μεταναστών ανά νομό για τα έτη 2007 και 2008, στα οποία περιέχονται ασφαλισμένοι του ΟΑΕΕ αλλά και του ΙΚΑ. Καθώς το αποτέλεσμα περιέχει γεωγραφική πληροφορία, τους νομούς της Ελλάδας, είναι δυνατόν να παρουσιαστεί και σε χάρτη GIS, όπως φαίνεται στο σχήμα 6-2.

	ΕΤΟΣ		
	2007	2008	Grand Total
DIM GEOGRAPHY	Metanasths Me Kwdikous Etwn Count	Metanasths Me Kwdikous Etwn Count	Metanasths Me Kwdikous Etwn Count
ΑΙΤΩΛΟΑΚΑΡΝΑΝΙΑΣ	700	695	724
ΑΡΓΟΛΙΔΑΣ	1289	1314	1329
ΑΡΚΑΔΙΑΣ	742	753	765
ΑΡΤΗΣ	171	186	193
ΑΤΤΙΚΗΣ	67250	68767	70312
ΑΧΑΪΔΑΣ	2770	3019	3105
ΒΟΙΩΤΙΑΣ	2623	2647	2664
ΓΡΕΒΕΝΩΝ	205	215	226
ΔΡΑΜΑΣ	946	992	1043
ΔΩΔΕΚΑΝΗΣΟΥ	2299	2431	2515
ΕΒΡΟΥ	1265	1258	1378
ΕΥΒΟΙΑΣ	1947	1965	2031
ΕΥΡΥΤΑΝΙΑΣ	56	56	61
ΖΑΚΥΝΘΟΥ	455	456	474
ΗΛΕΙΑΣ	611	628	650
ΗΜΑΘΙΑΣ	782	776	841
ΗΡΑΚΛΕΙΟΥ	1808	1883	1945
ΘΕΣΣΑΛΟΝΙΚΗΣ	373	377	401
ΘΕΣΣΑΛΟΝΙΚΗΣ	13263	13726	14378
ΙΩΑΝΝΙΝΩΝ	1215	1244	1299
ΚΑΒΑΛΑΣ	1278	1274	1367
ΚΑΡΔΙΤΣΑΣ	268	274	288
ΚΑΣΤΟΡΙΑΣ	300	305	326
ΚΕΡΚΥΡΑΣ	1043	1053	1104
ΚΕΦΑΛΛΗΝΙΑΣ	333	334	351
ΚΙΛΙΚΙΣ	621	634	673
ΚΟΖΑΝΗΣ	811	839	893
ΚΟΡΙΝΘΙΑΣ	2390	2412	2480
ΚΥΚΛΑΔΩΝ	1437	1496	1514
ΛΑΚΩΝΙΑΣ	679	686	702
ΛΑΡΙΣΣΗΣ	1374	1413	1464
ΛΑΣΙΘΙΟΥ	504	512	542
ΛΕΣΒΟΥ	741	751	778
ΛΕΥΚΑΔΑΣ	150	157	164
ΜΑΓΝΗΣΙΑΣ	1224	1284	1335
ΜΕΣΣΗΝΙΑΣ	1183	1219	1253
ΜΑΝΗΣ	692	705	758
ΠΕΛΛΗΝΗΣ	869	861	930
ΠΕΡΙΣΣΗΣ	1374	1429	1508
ΠΡΕΒΕΖΗΣ	430	438	474
ΡΕΘΥΜΝΗΣ	505	530	552
ΡΟΔΟΠΗΣ	614	655	705
ΣΑΜΟΥ	412	423	440
ΣΕΡΡΩΝ	282	293	308
ΤΡΙΚΑΛΩΝ	572	589	626
ΦΘΙΩΤΙΑΣ	937	952	973
ΦΛΩΡΙΝΑΣ	234	227	246
ΦΩΚΙΑΣ	200	204	207
ΧΑΛΚΙΔΙΚΗΣ	785	809	848
ΧΑΝΙΩΝ	1167	1215	1272
ΧΙΟΥ	398	398	416

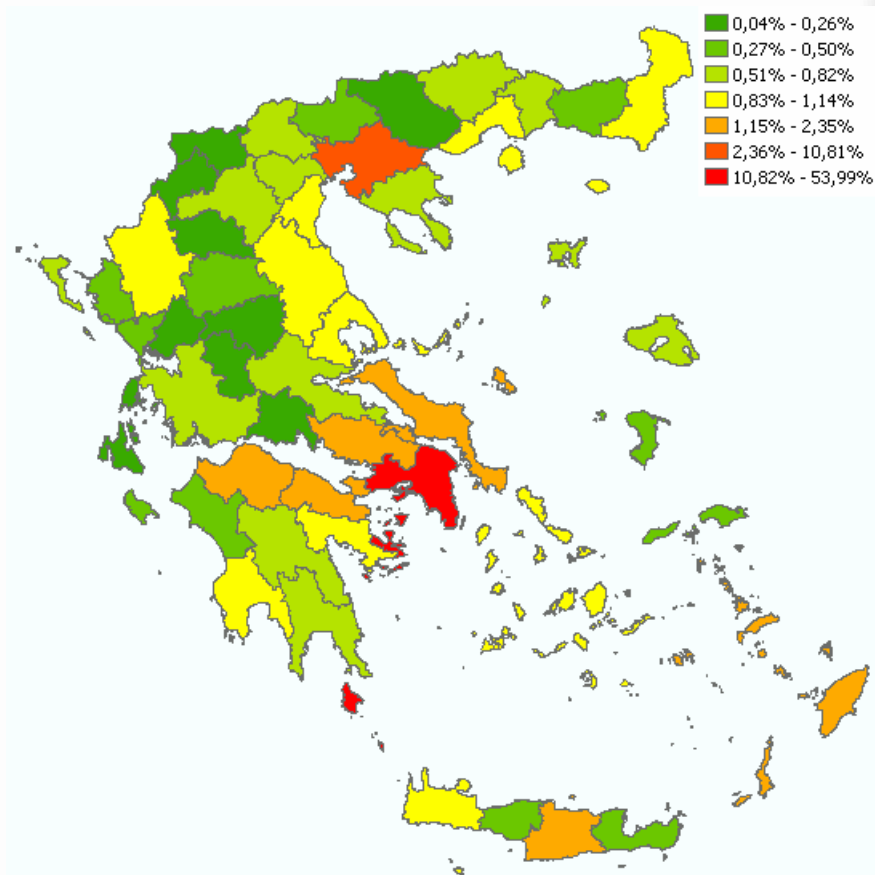
Σχ. 6-1 Πλήθος μεταναστών ανά νομό για 2008-09

Τα αποτελέσματα του σχήματος 6-1 θα οπτικοποιηθούν σε χάρτη, συνδέοντας τους νομούς του 6-1 με τους νομούς του χάρτη, οπότε σαν αποτέλεσμα έχουμε το σχήμα 6-2, στο οποίο φαίνονται πόσοι ασφαλισμένοι μετανάστες υπάρχουν σε κάθε νομό της Ελλάδας για το 2008. Ο χρωματισμός του κάθε νομού οφείλεται στο εύρος του πλήθους των μεταναστών. Δεξιά φαίνεται και ποιο χρώμα αντιστοιχεί σε ποιο εύρος.



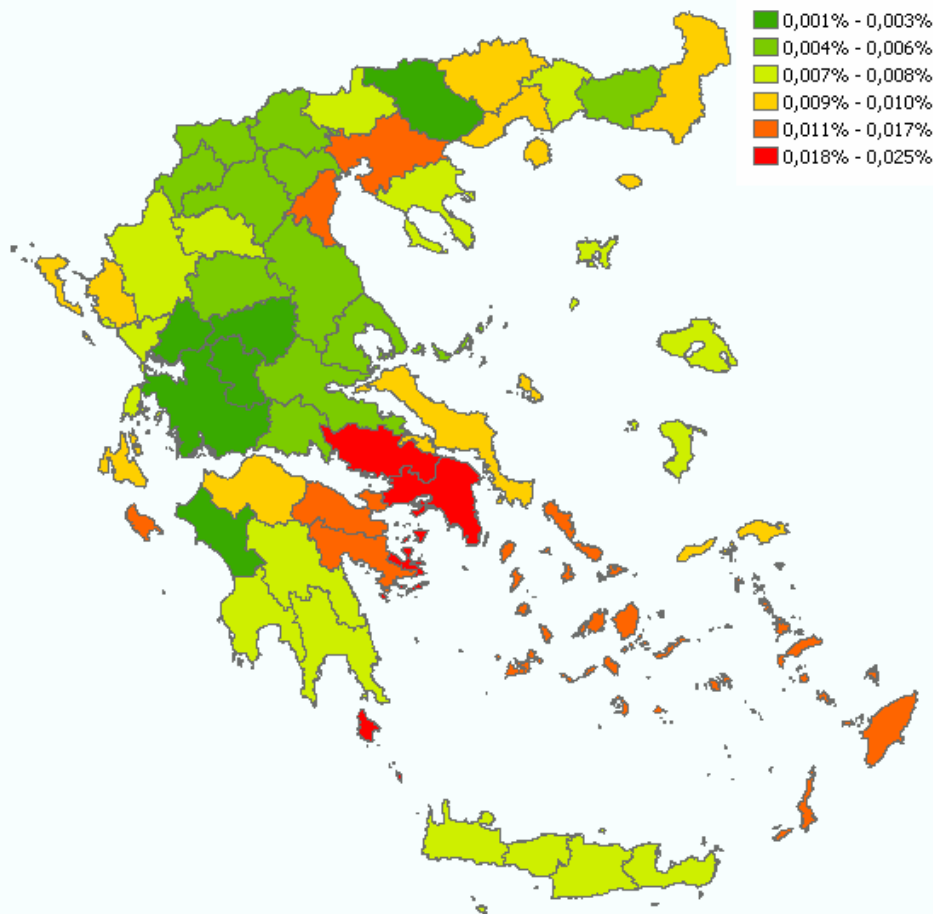
Σχ. 6-2 Πλήθος μεταναστών ανά νομό για το 2008

Παρατηρούμε από τους χρωματισμούς σε Αττική και Θεσσαλονίκη πως εκεί το πλήθος των μεταναστών είναι το μεγαλύτερο. Στο σχήμα 6-3 φαίνεται η ποσόστωση του πλήθους των μεταναστών σε κάθε νομό διά του συνόλου των μεταναστών της επικράτειας για το έτος 2008.



Σχ. 6-3 Ποσοστό μεταναστών ανά νομό ως προς το σύνολο των μεταναστών της Ελλάδος για το 2008

Προφανώς ανάμεσα στα σχήματα 6-2 και 6-3 δε φαίνονται ιδιαίτερες διαφορές, διαφορές όμως που θα είναι εμφανείς όταν υπολογιστεί το ποσοστό των μεταναστών ανά νομό ως προς το σύνολο των κατοίκων του νομού. Ο αριθμός των κατοίκων του κάθε νομού υπήρχε στα δεδομένα του αρχικού shapefile το οποίο εισήχθη στο χάρτη. Το αποτέλεσμα στο χάρτη φαίνεται στο σχήμα 6-4. Οι μεταβολές ανάμεσα στο σχήμα 6-3 και στο σχήμα 6-4 είναι εμφανείς. Παρατηρούμε ότι η Αττική εξακολουθεί να έχει τα πρωτεία ακόμα και στην ποσόστωση ως προς το σύνολο του πληθυσμού, ενώ νομοί όπως των Κυκλάδων ανεβαίνουν κλίμακα. Εν αντιθέσει νομοί όπως των Χανίων και του Ηρακλείου αν και στο σχήμα 6-3 είναι σε μία μέση κλίμακα, στο σχήμα 6-4 έχουν καταταχθεί σε χαμηλότερη. Η επιτυχία του χάρτη 6-4 είναι ότι από το χρωματισμό μπορούμε να καταλάβουμε ότι περιοχές οι οποίες βρίσκονται κοντά σε μεγάλα αστικά κέντρα (Αττική, Θεσσαλονίκη, Πάτρα) έχουν μεγάλο ποσοστό μεταναστών. Οι περιοχές που δεν ακολουθούν αυτόν τον κανόνα είναι κατά κύριο λόγο οι Κυκλάδες και τα Δωδεκάνησα. Από αυτή την παρατήρηση μπορούμε να οδηγηθούμε σε ένα νέο ερώτημα OLAP, το οποίο θα μας δώσει κάποια εξήγηση του φαινομένου και θα φανεί στην ενότητα 6.2



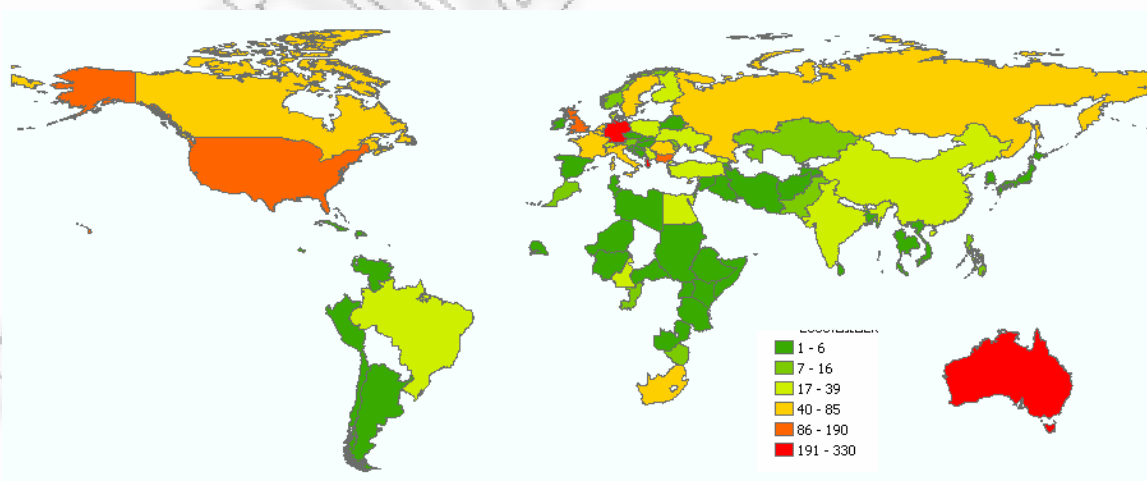
Σχ. 6-4 Ποσοστό μεταναστών ανά νομό ως προς το σύνολο των κατοίκων για το 2008

6.2 Πλήθος ατόμων ανά Χώρα για το έτος 2008 για τους νομούς Δωδεκανήσου και Κυκλάδων

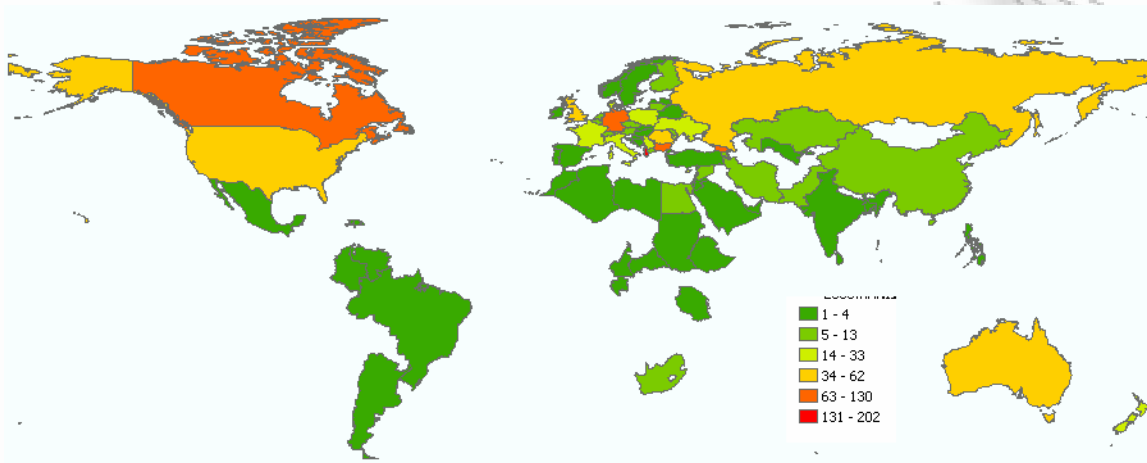
Από την τελευταία παρατήρηση της ενότητας 6.1 θα δημιουργήσουμε ένα νέο ερώτημα OLAP, στο οποίο θα παρουσιάζονται οι χώρες από τις οποίες προέρχονται οι ασφαλισμένοι μετανάστες για τους νομούς Κυκλάδων και Δωδεκανήσου, ώστε να προσπαθήσουμε να εξηγήσουμε τους λόγους της ύπαρξης μεγάλου πλήθους μεταναστών σε σύγκριση με τον πληθυσμό, παρόλο που οι δύο νομοί δεν είναι μεγάλα αστικά κέντρα. Στο σχήμα 6-5 βλέπουμε ένα μέρος των αποτελεσμάτων της OLAP ανάλυσης για τους 2 αυτούς νομούς. Για να προβούμε στην οπτικοποίηση σε χάρτη του αποτελέσματος θα χρησιμοποιήσουμε το χάρτη που περιέχει όλες τις χώρες, τον «xwres.mxd». Ενδεικτικά θα χρησιμοποιήσουμε το νομό Δωδεκανήσου από τους δύο αυτούς νομούς. Στο σχήμα 6-6 φαίνονται οι χώρες από τις οποίες προέρχονται οι μετανάστες του νομού αυτού χρωματισμένες ανάλογα με το πλήθος των μεταναστών. Θα χρησιμοποιήσουμε και έναν τρίτο νομό, το νομό Χανίων, ο οποίος έχει τα ίδια χαρακτηριστικά με τους άλλους δύο (τουριστικός), αλλά δεν παρουσιάζει το ίδιο φαινόμενο (μεγάλο ποσοστό μεταναστών) και φαίνεται στο σχήμα 6-7. Τελειώνοντας μπορούμε να κάνουμε σύγκριση ανάμεσα στα δύο σχήματα 6-6 και 6-7..

		ETOS	
		2008	Grand Total
DIM GEOGRAPHY	DESC ΧΩΡΑ ΚΑΤΑΓ	Metanasths Me Kwdikou	Metanasths Me Kwdikou
<input checked="" type="checkbox"/> ΚΥΚΛΑΔΩΝ		1496	1496
<input checked="" type="checkbox"/> ΔΩΔΕΚΑΝΗΣΟΥ	ΑΛΒΑΝΙΑ	330	330
	ΑΥΣΤΡΑΛΙΑ	289	289
	ΟΜΟΣΠ. ΔΗΜ. ΤΗΣ ΓΕΡΜΑΝΙΑΣ	280	280
	ΗΝΩΜΕΝΕΣ ΠΟΛ. ΤΗΣ ΑΜΕΡΙΚΗΣ	190	190
	ΗΝΩΜΕΝΟ ΒΑΣΙΛΕΙΟ	145	145
	ΒΟΥΛΓΑΡΙΑ	108	108
	ΙΤΑΛΙΑ	87	87
	ΚΑΝΑΔΑΣ	65	65
	ΣΟΥΗΔΙΑ	61	61
	ΝΟΤΙΑ ΑΦΡΙΚΗ	55	55
	ΚΑΤΩ ΧΩΡΕΣ	53	53
	ΡΩΣΙΑ	53	53
	ΒΕΛΓΙΟ	46	46
	ΡΟΥΜΑΝΙΑ	46	46
	ΓΑΛΛΙΑ	44	44
	ΚΥΠΡΟΣ	42	42
	ΑΙΓΥΠΤΟΣ	39	39
	ΟΥΚΡΑΝΙΑ	38	38
	ΔΑΝΙΑ	33	33
	ΣΕΡΒΙΑ	28	28
	ΦΙΝΛΑΝΔΙΑ	27	27
DIM GEOGRAPHY	DESC ΧΩΡΑ ΚΑΤΑΓ	2008	Grand Total
<input checked="" type="checkbox"/> ΚΥΚΛΑΔΩΝ		494	494
	ΑΛΒΑΝΙΑ	114	114
	ΒΟΥΛΓΑΡΙΑ	102	102
	ΡΩΣΙΑ	82	82
	ΟΜΟΣΠ. ΔΗΜ. ΤΗΣ ΓΕΡΜΑΝΙΑΣ	57	57
	ΡΟΥΜΑΝΙΑ	55	55
	ΠΟΛΩΝΙΑ	52	52
	ΙΤΑΛΙΑ	51	51
	ΗΝΩΜΕΝΟ ΒΑΣΙΛΕΙΟ	47	47
	ΓΑΛΛΙΑ	43	43
	ΗΝΩΜΕΝΕΣ ΠΟΛ. ΤΗΣ ΑΜΕΡΙΚΗΣ	35	35
	ΚΑΝΑΔΑΣ	30	30
	ΑΙΓΥΠΤΟΣ	28	28
	ΑΥΣΤΡΑΛΙΑ	27	27
	ΚΥΠΡΟΣ	27	27
	ΟΥΚΡΑΝΙΑ	27	27
	ΣΕΡΒΙΑ	22	22

Σχ. 6-5 Σύνολο μεταναστών για Κυκλάδες και Δωδεκάνησα ανά χώρα για το 2008



Σχ. 6-6 Σύνολο μεταναστών ανά χώρα για το 2008 για τα Δωδεκάνησα



Σχ. 6-7 Σύνολο μεταναστών ανά χώρα για το 2008 για το νομό Χανίων

Από τους δύο τελευταίους χάρτες δε μπορούμε να βγάλουμε κάποιο χαρακτηριστικό συμπέρασμα, τόσο λόγω του πλήθους των χωρών που καθιστά δυσδιάκριτη κάποια λεπτομέρεια, όσο και λόγω της μη ύπαρξης κάποιου ιδιαίτερου χαρακτηριστικού.

6.3 Πλήθος ατόμων ανά Έτος και ανά Νομό διαμονής και ανά Φύλο

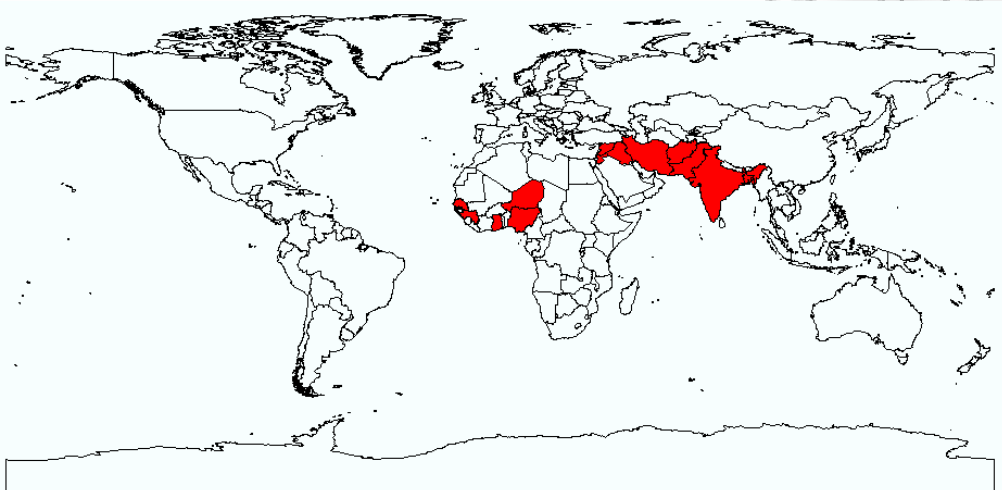
Σε αυτή την ενότητα θα αναζητήσουμε το πλήθος των μεταναστών ανά φύλο και ανά χώρα για το έτος 2008. Ένα μέρος των αποτελεσμάτων φαίνεται στο σχήμα 6-8. Δεν υπάρχει λόγος να οπτικοποιήσουμε το αποτέλεσμα σε χάρτη, καθώς όπως είδαμε από το προηγούμενο παράδειγμα είναι δυσδιάκριτη η εξαγωγή συμπεράσματος σε χάρτη που περιέχει όλες τις χώρες.

DESC ΧΩΡΑ ΚΑΤΑΓ	ETOS FYLO		Total	Grand Total
	A	Γ		
ALBANIA	65,00%	35,00%	100,00%	100
OMOSP. DHM. THS GERMANIAS	63,23%	36,77%	100,00%	100
PAKISTAN	99,69%	0,31%	100,00%	100
ROSIA	41,69%	58,31%	100,00%	100
INDIA	97,44%	2,56%	100,00%	100
ROYMANIA	49,51%	50,49%	100,00%	100
GEORGIA	45,66%	54,34%	100,00%	100
AUSTRALIA	62,03%	37,97%	100,00%	100
BOYLGARIA	35,78%	64,22%	100,00%	100
AIGYPTOS	79,07%	20,93%	100,00%	100
TOYRKIA	69,75%	30,25%	100,00%	100
HNOMENES POL. THS AMERIKHS	54,42%	45,58%	100,00%	100
MΠΑΓΚΛΑΝΤΕΣ	99,13%	0,87%	100,00%	100
KYΠΡΟΣ	60,90%	39,10%	100,00%	100
KANADAZ	60,75%	39,25%	100,00%	100
HNOMENO BASILEIO	49,85%	50,15%	100,00%	100
KAZAKSTAN	56,34%	43,66%	100,00%	100
ZYRIA	93,44%	6,56%	100,00%	100
POLONIA	45,74%	54,26%	100,00%	100
ITALIA	60,25%	39,75%	100,00%	100
KINA	65,89%	34,11%	100,00%	100
GALIA	53,79%	46,21%	100,00%	100
ARMENIA	51,67%	48,33%	100,00%	100
NOTIA AFRIKH	58,01%	41,99%	100,00%	100
SOYNDIA	57,88%	42,12%	100,00%	100
IRAK	94,11%	5,89%	100,00%	100
OYKRAVIA	21,96%	78,04%	100,00%	100
MIPHRIA	86,87%	13,13%	100,00%	100
BEΛΓIO	56,86%	41,14%	100,00%	100
SERBIA				

Σχ. 6-8 Ποσοστό επί του φύλου μεταναστών ανά χώρα για το 2008

Στο σχήμα 6-8 βλέπουμε την ποσόστωση του κάθε φύλλου στο σύνολο των μεταναστών της χώρα καταγωγής τους. Η Αλβανία π.χ. έχει 65% ανδρικό πληθυσμό και 35% θήλυ. Σε αυτό το κομμάτι των αποτελεσμάτων, ιδιαίτερα εντυπωσιακό είναι τα ποσοστά των μουσουλμανικών

χωρών. Βλέπουμε ότι χώρες όπως το Πακιστάν, το Μπαγκλαντές, το Ιράκ αλλά και άλλες όπως η Ινδία και η Συρία έχουν πολύ μεγάλα ποσοστά, της τάξεως άνω του 90%, παρουσίας ανδρικού πληθυσμού. Ενδιαφέρον θα ήταν να παρουσιαστούν σε χάρτη οι χώρες οι οποίες έχουν ανδρικό πληθυσμό μεταναστών στην Ελλάδα, άνω του 80%. Αυτό μπορούμε να το δούμε στο σχήμα 6-9.



Σχ. 6-9 Χώρες με ποσοστό ανδρικού πληθυσμού μεταναστών άνω του 80% για το 2008

6.4 Πλήθος ατόμων ανά Έτος και ανά Ήπειρο καταγωγής και ανά Οικογενειακή κατάσταση

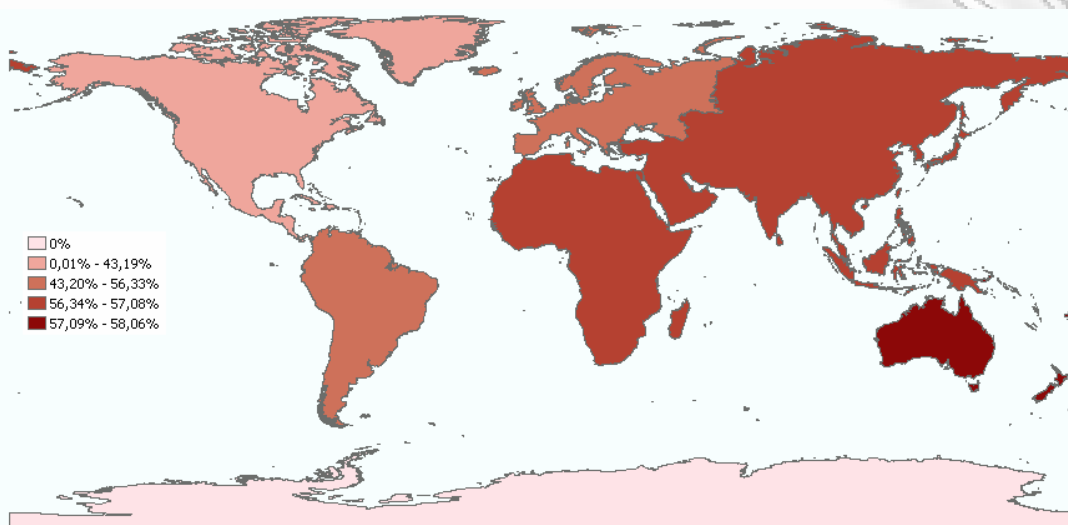
Καθώς δεδομένα οικογενειακής κατάστασης υπάρχουν μόνο για τους ασφαλισμένους μετανάστες του ΟΑΕΕ, σε αυτή την ενότητα θα εξετάσουμε την οικογενειακή κατάσταση ανά ήπειρο για τους ασφαλισμένους του ΟΑΕΕ για τα έτη 2007-2009. Στο σχήμα 6-10 φαίνεται το ποσοστό των ασφαλισμένων που είναι έγγαμοι ή όχι, ανά ήπειρο για το 2009.

		ETOS ▾ DESC ΟΙΚΟΓ ΚΑΤ. ▾			
		☐ 2009		Grand Total	
		ΑΓΑΜΟΣ	ΕΓΓΑΜΟΣ	Total	
ASFAL FO ▾	ΗΠΕΙΡΟΣ ▾	Metanasths Me Kwc	Metanasths Me Kwc	Metanasths Me Kwc	Metanasths Me Kwc
☐ ΟΑΕΕ	AF	43,44%	56,56%	100,00%	100,00%
	AS	42,92%	57,08%	100,00%	100,00%
	AU	42,12%	57,88%	100,00%	100,00%
	EU	44,11%	55,89%	100,00%	100,00%
	NA	56,81%	43,19%	100,00%	100,00%
	OC	41,94%	58,06%	100,00%	100,00%
	SA	43,67%	56,33%	100,00%	100,00%
	Total	44,74%	55,26%	100,00%	100,00%
Grand Total		44,74%	55,26%	100,00%	100,00%

Σχ. 6-10 Ποσοστό έγγαμων και άγαμων ασφαλισμένων του ΟΑΕΕ ανά ήπειρο για το 2009

Παρατηρούμε ότι τα ποσοστά έγγαμων με άγαμους δεν έχουν ουσιαστική διαφορά ανά ήπειρο. Σε όλες τις ηπείρους το ποσοστό των έγγαμων είναι ανάμεσα στο 55% με 58%. Η μόνη εμφανής διαφορά είναι για τη Βόρεια Αμερική όπου το ποσοστό αντιστρέφεται. Στη συγκεκριμένη ήπειρο το ποσοστό των άγαμων μεταναστών είναι στο 56,8%.

Στο σχήμα 6-11 βλέπουμε τον αντίστοιχο χάρτη για τις ηπείρους με χρωματισμό της προσότωσης των έγγαμων ασφαλισμένων ως προς το σύνολο των μεταναστών ασφαλισμένων κάθε ηπείρου.



Σχ. 6-11 Ήπειροι με ποσοστό έγγαμων ασφαλισμένων του ΟΑΕΕ για το 2009

6.5 Πλήθος ατόμων ανά Επαγγελματική δραστηριότητα και ανά Έτος

Στη συγκεκριμένη ενότητα θα αναλύσουμε το πλήθος των ατόμων ανά οικονομική δραστηριότητα, ώστε να δούμε οι μετανάστες που έρχονται στην Ελλάδα σε τι τομείς απασχολούνται για τα έτη 2004 έως 2008. Καθώς έχουμε δεδομένα μόνο για το ΙΚΑ η ανάλυση θα γίνει μόνο για το συγκεκριμένο φορέα.

ASFAΛ FOREAS ΙΚΑ ΕΤΑΜ	ΕΤΟΣ				
	2004	2005	2006	2007	2008
DESC ΟΙΚΟΝΟΜ ΔΡΑΣΤ	Metanasths	Metanasths	Metanasths	Metanasths	Metanasths
ΜΕΤΑΠΟΙΗΤΙΚΕΣ ΒΙΟΜΗΧΑΝΙΕΣ	28,90%	27,84%	27,70%	30,49%	27,94%
ΞΕΝΟΔΟΧΕΙΑ ΚΑΙ ΕΣΤΙΑΤΟΡΙΑ	20,95%	21,50%	24,02%	24,68%	22,54%
ΧΟΝΔΡΙΚΟ ΚΑΙ ΛΙΑΝΙΚΟ ΕΜΠΟΡΙΟ - ΕΠΙΣΚΕΥΗ ΑΥΤΟΚΙΝΗΤΩΝ ΟΧΗΜΑΤΩΝ, ΜΟΤΟΣΥΚΛΙ ΚΑΤΑΣΚΕΥΕΣ	19,42%	19,74%	20,32%	21,16%	20,21%
ΙΔΙΩΤΙΚΑ ΝΟΙΚΟΚΥΡΙΑ ΠΟΥ ΑΠΑΣΧΟΛΟΥΝ ΟΙΚΙΑΚΟ ΠΡΟΣΩΠΙΚΟ ΚΑΙ ΜΗ ΔΙΑΦΟΡΟΠΟΙ	11,68%	11,27%	10,36%	6,87%	5,73%
ΔΙΑΧΕΙΡΙΣΗ ΑΚΙΝΗΤΗΣ ΠΕΡΙΟΥΣΙΑΣ, ΕΚΜΙΣΘΩΣΕΙΣ ΚΑΙ ΕΠΙΧΕΙΡΗΜΑΤΙΚΕΣ ΔΡΑΣΤΗΡΙΟΤ	7,95%	6,82%	6,43%	7,38%	5,99%
ΜΕΤΑΦΟΡΕΣ, ΑΠΟΘΗΚΕΥΣΗ ΚΑΙ ΕΠΙΚΟΙΝΩΝΙΕΣ	6,94%	7,09%	7,20%	6,79%	6,70%
ΆΛΛΕΣ ΔΡΑΣΤΗΡΙΟΤΗΤΕΣ ΠΑΡΟΧΗΣ ΥΠΗΡΕΣΙΩΝ ΥΠΕΡ ΤΟΥ ΚΟΙΝΩΝΙΚΟΥ ΣΥΝΟΛΟΥ ΚΑ	3,83%	4,01%	3,93%	3,49%	3,53%
ΕΚΠΑΙΔΕΥΣΗ	3,49%	3,68%	3,53%	3,42%	3,32%
ΥΓΕΙΑ ΚΑΙ ΚΟΙΝΩΝΙΚΗ ΜΕΡΙΜΝΑ	2,83%	3,01%	2,76%	1,61%	1,09%
ΓΕΩΡΓΙΑ, ΚΤΗΝΟΤΡΟΦΙΑ, ΘΗΡΑ ΚΑΙ ΔΑΣΟΚΟΜΙΑ	1,77%	1,76%	1,65%	1,29%	1,18%
ΕΝΔΙΑΜΕΣΟΙ ΧΡΗΜΑΤΟΠΙΣΤΩΤΙΚΟΙ ΟΡΓΑΝΙΣΜΟΙ	0,74%	0,54%	0,55%	0,55%	0,52%
ΟΡΥΧΕΙΑ ΚΑΙ ΛΑΤΟΜΕΙΑ	0,73%	0,71%	0,62%	0,35%	0,30%
ΑΛΙΕΙΑ	0,44%	0,49%	0,50%	0,49%	0,43%
ΔΗΜΟΣΙΑ ΔΙΟΙΚΗΣΗ ΚΑΙ ΑΜΥΝΑ - ΥΠΟΧΡΕΩΤΙΚΗ ΚΟΙΝΩΝΙΚΗ ΑΣΦΑΛΙΣΗ	0,30%	0,21%	0,23%	0,26%	0,21%
ΠΑΡΟΧΗ ΗΛΕΚΤΡΙΚΟΥ ΡΕΥΜΑΤΟΣ, ΦΥΣΙΚΟΥ ΑΕΡΙΟΥ ΚΑΙ ΝΕΡΟΥ	0,17%	0,18%	0,20%	0,18%	0,18%
ΕΤΕΡΟΔΙΚΟΙ ΟΡΓΑΝΙΣΜΟΙ ΚΑΙ ΟΡΓΑΝΑ	0,10%	0,10%	0,12%	0,11%	0,12%
Grand Total	0,08%	0,11%	0,14%	0,06%	0,04%
	100,00%	100,00%	100,00%	100,00%	100,00%

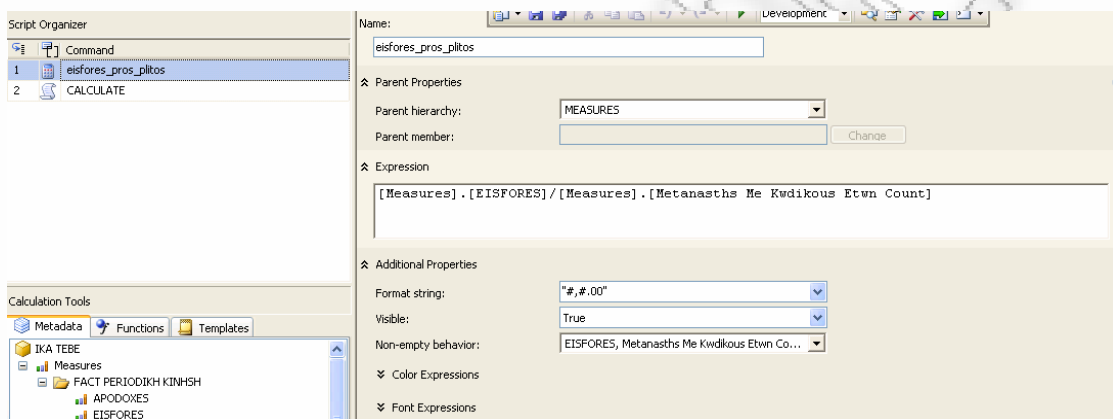
Σχ. 6-12 Είδος επαγγελματικής δραστηριότητας ανά έτος

Παρατηρούμε ότι η πλειονότητα των μεταναστών απασχολείται σε «Μεταποιητικές Βιομηχανίες», «Ξενοδοχεία και Εστιατόρια», «Επισκευές οχημάτων» και «Κατασκευές». Ενώ οι τρεις πρώτες κατηγορίες έχουν μια αυξομειούμενη αλλά σταθερή προσότωση, η τελευταία κατηγορία έχει μια εντυπωσιακή μείωση της τάξεως του 50%. Από το 2004 έως το 2008 οι μετανάστες που απασχολούνταν στις «Κατασκευές» μειώθηκαν από το 11,68% στο 5,73% του συνόλου των μεταναστών, γεγονός το οποίο είναι κατανοητό καθώς το 2004 ήταν η περίοδος

των Ολυμπιακών Αγώνων με έντονη κατασκευαστική δραστηριότητα, ενώ παράλληλα και η οικονομική κρίση έναν από τους τομείς που επηρέασε πλήρως ήταν αυτός των κατασκευών.

6.6 Ποσό εισφορών ανά χώρα καταγωγής

Στην ενότητα 6.6 θα αναλύσουμε το σύνολο των εισφορών που δίνουν οι μετανάστες ανά χώρα καταγωγής. Μέσα από αυτή την ανάλυση μπορούμε να εντοπίσουμε ποιες χώρες «προσφέρουν περισσότερο» στο ασφαλιστικό μας σύστημα και να προσπαθήσουμε να προσελκύσουμε περισσότερα άτομα από κείνες τις περιοχές. Καθώς δεδομένα εισφορών υπάρχουν μόνο για τον ΟΑΕΕ, το όποιο αποτέλεσμα θα αφορά μόνο αυτό το φορέα, ενώ θα επικεντρωθούμε σε ένα συγκεκριμένο έτος, το 2009. Για να είναι το αποτέλεσμα πιο λογικό, θα πρέπει να εμφανιστεί το ποσό των εισφορών ανά μετανάστη. Για να υλοποιηθεί αυτό, πρέπει να δημιουργηθεί ένα νέο μέτρο το «Σύνολο εισφορών ανά μετανάστη». Στο σχήμα 6-13 παρουσιάζεται πως δημιουργείται το μέτρο στο Analysis Services.



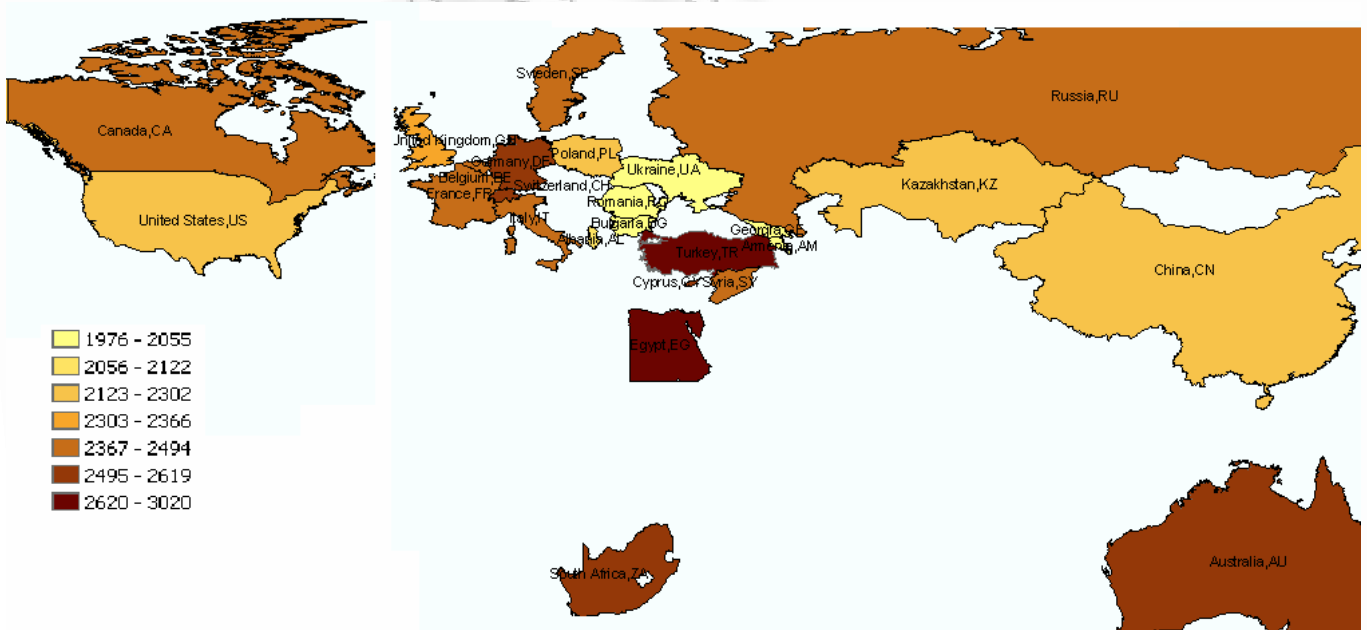
Σχ. 6-13 Δημιουργία νέου μέτρου «Σύνολο εισφορών ανά μετανάστη»

Στο σχήμα 6-14 φαίνονται οι χώρες των οποίων οι πολίτες που έχουν μεταναστεύσει στην Ελλάδα πληρώνουν τις περισσότερες εισφορές σε ευρώ ανά μετανάστη. Καθώς υπάρχουν χώρες που έχουν μόνο ένα μετανάστη, επιλέγονται οι 25 χώρες με τους περισσότερους μετανάστες για πιο ασφαλή συμπεράσματα.

ASFAL FOREAS		ETOS			
ΟΑΕΕ		2009		Grand Total	
DESC ΧΩΡΑ ΚΑΤΑΓ	Metanasths	eisfores_pros_plitos	Metanasths	eisfores_pros_plitos	
ΤΟΥΡΚΙΑ	2085	3.020,28	2085	3.020,28	
ΑΙΓΥΠΤΟΣ	1489	2.854,58	1489	2.854,58	
ΑΥΣΤΡΑΛΙΑ	3261	2.619,38	3261	2.619,38	
ΕΛΒΕΤΙΑ	479	2.539,37	479	2.539,37	
ΚΥΠΡΟΣ	1609	2.537,88	1609	2.537,88	
ΝΟΤΙΑ ΑΦΡΙΚΗ	818	2.523,37	818	2.523,37	
ΟΜΟΣΠ. ΔΗΜ. ΤΗΣ ΓΕΡΜΑΝΙΑΣ	14545	2.522,14	14545	2.522,14	
ΒΕΛΓΙΟ	665	2.494,08	665	2.494,08	
ΚΑΝΑΔΑΣ	2068	2.474,61	2068	2.474,61	
ΙΤΑΛΙΑ	1019	2.468,05	1019	2.468,05	
ΓΑΛΛΙΑ	827	2.444,80	827	2.444,80	
ΡΩΣΙΑ	2973	2.442,90	2973	2.442,90	
ΣΟΥΗΔΙΑ	816	2.436,30	816	2.436,30	
ΣΥΡΙΑ	631	2.432,26	631	2.432,26	
ΗΝΩΜΕΝΟ ΒΑΣΙΛΕΙΟ	1779	2.365,64	1779	2.365,64	
ΗΝΩΜΕΝΕΣ ΠΟΛ. ΤΗΣ ΑΜΕΡΙΚΗΣ	2694	2.302,15	2694	2.302,15	
ΠΟΛΩΝΙΑ	917	2.284,66	917	2.284,66	
ΚΑΖΑΚΣΤΑΝ	1419	2.266,37	1419	2.266,37	
ΚΙΝΑ	773	2.261,64	773	2.261,64	
ΑΛΒΑΝΙΑ	8023	2.121,72	8023	2.121,72	
ΓΕΩΡΓΙΑ	2599	2.054,97	2599	2.054,97	
ΑΡΜΕΝΙΑ	582	2.049,40	582	2.049,40	
ΡΟΥΜΑΝΙΑ	1444	2.031,39	1444	2.031,39	
ΟΥΚΡΑΝΙΑ	661	2.003,29	661	2.003,29	
ΒΟΥΛΓΑΡΙΑ	1416	1.976,49	1416	1.976,49	
Grand Total	55592	2.398,16	55592	2.398,16	

Σχ. 6-14 Εισφορές ανά μετανάστη για τις 25 χώρες με τους περισσότερους μετανάστες

Ακολουθως στο σχήμα 6-15 παρουσιάζεται ο χάρτης με τις χώρες χρωματισμένες ανάλογα με συγκεκριμένο εύρος πλήθους εισφορών (σε ευρώ) ανά μετανάστη.



Σχ. 6-15 Οι 25 χώρες με τους περισσότερους μετανάστες χρωματισμένες ανάλογα με το πλήθος των εισφορών ανά μετανάστη

Παρατηρούμε ότι οι μετανάστες από ευρωπαϊκές αλλά και από τις μεγάλες εδαφικά χώρες είναι που ωφελούν περισσότερο το ελληνικό ασφαλιστικό σύστημα. Χαρακτηριστική είναι η απουσία όλων των χωρών της Νότιας Αμερικής από τον χάρτη των 25 αυτών χωρών με τους περισσότερους μετανάστες.

6.7 Πλήθος ατόμων ανά χώρα καταγωγής και ανά Οικονομική δραστηριότητα

Στη συγκεκριμένη ενότητα θα παρουσιαστεί ποια οικονομική δραστηριότητα ασκούν ανάλογα με τη χώρα καταγωγής τους οι πολίτες που έχουν μεταναστεύσει στην Ελλάδα. Όπως αναφέρθηκε και νωρίτερα, καθώς δεν υπάρχουν στοιχεία απασχόλησης για τους ασφαλισμένους του ΟΑΕΕ, θα επικεντρωθούμε μόνο σε αυτούς του ΙΚΑ για το έτος 2008 και για πέντε συγκεκριμένες κατηγορίες, οι οποίες φαίνονται στο σχήμα 6.16. Λόγω του μεγάλου πλήθους των χωρών στο σχήμα 6-16 φαίνονται οι 5 χώρες με τους περισσότερους μετανάστες στην Ελλάδα ανά την κάθε οικονομική δραστηριότητα.

ASFAL FOREAS ▾ ETOS ▾			
ΙΚΑ ΕΤΑΜ		2008	
DESC ΟΙΚΟΝΟΜ ΔΡΑΣΤ	DESC ΧΩΡΑ ΚΑΤΑΓ	Drop Col	Metanast
<input type="checkbox"/> ΜΕΤΑΠΟΙΗΤΙΚΕΣ ΒΙΟΜΗΧΑΝΙΕΣ	ΑΛΒΑΝΙΑ		7834
	ΠΑΚΙΣΤΑΝ		3067
	ΡΟΥΜΑΝΙΑ		1232
	ΡΩΣΙΑ		863
	ΒΟΥΛΓΑΡΙΑ		686
	Total		13682
<input type="checkbox"/> ΞΕΝΟΔΟΧΕΙΑ ΚΑΙ ΕΣΤΙΑΤΟΡΙΑ	ΑΛΒΑΝΙΑ		6948
	ΒΟΥΛΓΑΡΙΑ		1445
	ΡΟΥΜΑΝΙΑ		1157
	ΡΩΣΙΑ		1067
	ΠΑΚΙΣΤΑΝ		367
	Total		10984
<input type="checkbox"/> ΧΟΝΔΡΙΚΟ ΚΑΙ ΛΙΑΝΙΚΟ ΕΜΠΟΡΙΟ - ΕΠΙΣΚΕΥΗ ΑΥΤΟΚΙΝΗΤΩΝ ΟΧΗΜΑΤΩΝ, ΜΟΤΟΣΥ	ΑΛΒΑΝΙΑ		6412
	ΠΑΚΙΣΤΑΝ		1231
	ΒΟΥΛΓΑΡΙΑ		653
	ΡΩΣΙΑ		620
	ΡΟΥΜΑΝΙΑ		619
	Total		9535
<input type="checkbox"/> ΚΑΤΑΣΚΕΥΕΣ	ΑΛΒΑΝΙΑ		2285
	ΡΟΥΜΑΝΙΑ		255
	ΠΑΚΙΣΤΑΝ		233
	ΒΟΥΛΓΑΡΙΑ		153
	ΡΩΣΙΑ		143
	Total		3069
<input type="checkbox"/> ΔΙΑΧΕΙΡΙΣΗ ΑΚΙΝΗΤΗΣ ΠΕΡΙΟΥΣΙΑΣ, ΕΚΜΙΣΘΩΣΕΙΣ ΚΑΙ ΕΠΙΧΕΙΡΗΜΑΤΙΚΕΣ ΔΡΑΣΤΗΡ	ΑΛΒΑΝΙΑ		1961
	ΒΟΥΛΓΑΡΙΑ		410
	ΡΩΣΙΑ		306
	ΡΟΥΜΑΝΙΑ		223
	ΠΑΚΙΣΤΑΝ		155
	Total		3055

Σχ. 6-16 Πιο πολυπληθείς χώρες ανά οικονομική δραστηριότητα

Από το παραπάνω σχήμα παρατηρούμε την έντονη παρουσία των Αλβανών σε κάθε οικονομική δραστηριότητα, γεγονός το οποίο συνεχίζεται και για τις υπόλοιπες δραστηριότητες, οι οποίες δε φαίνονται στο σχήμα 6-16. Το πλήθος των Αλβανών είναι τόσο μεγάλο όπου η διαφορά από τις υπόλοιπες εθνικότητες είναι εντυπωσιακή.

6.8 Πλήθος ατόμων ανά φορέα και ανά χώρα γέννησης

Παρατηρώντας τις ενότητες 6.6 και 6.7 γεννάται η απορία ποιες είναι οι κυρίαρχες εθνικότητες ανά φορέα. Ο ΟΑΕΕ είναι ένας φορέας όπου ασφαρίζονται ελεύθεροι επαγγελματίες, ενώ το ΙΚΑ ασφαρίζει εργατικό δυναμικό. Στο σχήμα 6-17 μπορούμε να δούμε αυτή την ανάλυση για το έτος 2008, το οποίο περιέχει δεδομένα τόσο για τον ΟΑΕΕ όσο και για το ΙΚΑ.

ΕΤΟΣ ▾		Metanasths Me
2008		
ASFAL FOREAS ▾	DESC ΧΩΡΑ ΚΑΤΑΓ ▾	
☐ ΙΚΑ ΕΤΑΜ	ΑΛΒΑΝΙΑ	44,94%
	ΠΑΚΙΣΤΑΝ	8,34%
	ΒΟΥΛΓΑΡΙΑ	6,60%
	ΡΩΣΙΑ	5,98%
	ΡΟΥΜΑΝΙΑ	5,89%
	ΙΝΔΙΑ	4,20%
	ΓΕΩΡΓΙΑ	3,03%
	ΟΥΚΡΑΝΙΑ	2,40%
	ΜΠΑΓΚΛΑΝΤΕΣ	1,96%
	ΦΙΛΙΠΠΙΝΕΣ	1,55%
	ΠΟΛΩΝΙΑ	1,52%
ΑΙΓΥΠΤΟΣ	1,38%	
☐ ΟΑΕΕ	ΟΜΟΣΠ. ΔΗΜ. ΤΗΣ ΓΕΡΜΑΝΙΑΣ	23,65%
	ΑΛΒΑΝΙΑ	12,33%
	ΑΥΣΤΡΑΛΙΑ	5,26%
	ΡΩΣΙΑ	4,94%
	ΗΝΩΜΕΝΕΣ ΠΟΛ. ΤΗΣ ΑΜΕΡΙΚΗΣ	4,32%
	ΓΕΩΡΓΙΑ	4,11%
	ΤΟΥΡΚΙΑ	3,50%
	ΚΑΝΑΔΑΣ	3,32%
	ΗΝΩΜΕΝΟ ΒΑΣΙΛΕΙΟ	2,86%
	ΚΥΠΡΟΣ	2,62%
	ΑΙΓΥΠΤΟΣ	2,51%
	ΚΑΖΑΚΣΤΑΝ	2,34%

Σχ. 6-17 Πιο πολυπληθείς εθνικότητες ανά φορέα

Στο σχήμα 6-17 βλέπουμε ότι το 44,94 των ασφαλισμένων μεταναστών του ΙΚΑ είναι από την Αλβανία, ενώ οι υπόλοιπες εθνικότητες ακολουθούν με ποσοστό κάτω του 10%. Αντίθετα στον ΟΑΕΕ κυρίαρχη εθνικότητα είναι οι γεννηθέντες στη Γερμανία, ενώ ακολουθούν οι της Αλβανίας με ποσοστό της τάξεως του 12%. Παρατηρούμε ότι και στους δύο φορείς οι Αλβανοί έχουν έντονη παρουσία, είτε με τεράστια ποσοστά όπως στο ΙΚΑ, είτε καταλαμβάνοντας τη δεύτερη θέση στη σχετική λίστα του ΟΑΕΕ.

6.9 Πλήθος ατόμων ανά δεκαετία γέννησης

Στο συγκεκριμένο παράδειγμα θα εξεταστεί το πλήθος των μεταναστών για τον ΟΑΕΕ για το 2009 ανά δεκαετία γέννησης, το οποίο φαίνεται στο σχήμα 6-18. Παρατηρούμε μία αρκετά καλή αναλογία αν αναλογιστούμε ότι άτομα γεννημένα τη δεκαετία 1970-1979 είναι άτομα που βρίσκονται περίπου μεταξύ 30 και 40 ετών και είναι στην πλέον παραγωγική ηλικία. Το αρνητικό είναι η απουσία μεταναστών γεννημένων τη δεκαετία 1980-1989 (10%), οπότε ο ΟΑΕΕ θα έπρεπε να στοχεύει στην προσέλκυση μεταναστών νεαρής ηλικίας.

ETOS	ASFAL FOREAS
2009	ΟΑΕΕ
Drop Column Fie	
Dekaetia	Metanasths Me
1970-1979	40,73%
1960-1969	32,33%
1950-1959	13,31%
1980-1989	9,75%
1940-1949	3,70%
1930-1939	0,10%
1990-1999	0,09%
Grand Total	100,00%

Σχ. 6-18 Ποσοστό ασφαλισμένων μεταναστών ΟΑΕΕ για το 2009 ανά δεκαετία γέννησης

6.10 Πλήθος ατόμων ανά χώρα καταγωγής και ανά Δεκαετία γέννησης.

Στην ενότητα 6.10 αναλύεται το πλήθος των μεταναστών του ΟΑΕΕ για το 2009 ανά δεκαετία γέννησης και ανά χώρα γέννησης. Στο σχήμα 6-19 μπορούμε να δούμε τις πέντε πρώτες χώρες σε αριθμό μεταναστών στην Ελλάδα, για τις δεκαετίες '60, '70, '80. Παρατηρείται η έντονη παρουσία μεταναστών από τη Γερμανία γεννημένων το '60 και το '70, ενώ όσον αφορά τη δεκαετία του '80, παρατηρείται ότι οι γεννηθέντες στην Αλβανία είναι πλέον η κυρίαρχη κατηγορία μεταναστών στον ΟΑΕΕ.

ETOS	ASFAL FOREAS	
2009	ΟΑΕΕ	
Drop Column Fie		
Dekaetia	DESC XWRA KATAG	Metanasths Me
<input type="checkbox"/> 1960-1969	ΟΜΟΣΠ. ΔΗΜ. ΤΗΣ ΓΕΡΜΑΝΙΑΣ	5171
	ΑΛΒΑΝΙΑ	2294
	ΑΥΣΤΡΑΛΙΑ	1512
	ΡΩΣΙΑ	905
	ΗΝΩΜΕΝΕΣ ΠΟΛ. ΤΗΣ ΑΜΕΡΙΚΗΣ	599
	Total	10481
<input type="checkbox"/> 1970-1979	ΟΜΟΣΠ. ΔΗΜ. ΤΗΣ ΓΕΡΜΑΝΙΑΣ	8071
	ΑΛΒΑΝΙΑ	3169
	ΗΝΩΜΕΝΕΣ ΠΟΛ. ΤΗΣ ΑΜΕΡΙΚΗΣ	1451
	ΑΥΣΤΡΑΛΙΑ	1440
	ΡΩΣΙΑ	879
	Total	15010
<input type="checkbox"/> 1980-1989	ΑΛΒΑΝΙΑ	1365
	ΟΜΟΣΠ. ΔΗΜ. ΤΗΣ ΓΕΡΜΑΝΙΑΣ	828
	ΗΝΩΜΕΝΕΣ ΠΟΛ. ΤΗΣ ΑΜΕΡΙΚΗΣ	405
	ΡΩΣΙΑ	333
	ΑΥΣΤΡΑΛΙΑ	163
	Total	3094
Grand Total		28585

Σχ. 6-19 Πλήθος μεταναστών ανά δεκαετία γέννησης και χώρα καταγωγής

6.11 Πλήθος ατόμων ανά Δεκαετία εγγραφής

Πολύ ενδιαφέρον θα είχε να αναλυθεί το πλήθος των μεταναστών ανά δεκαετία κατά την οποία εγγράφηκαν σε έναν ασφαλιστικό φορέα. Θα γίνει ανάλυση αυτού του δεδομένου για τους μετανάστες που ήταν ενεργοί ασφαλισμένοι κατά το 2008. Στο σχήμα 6-20 βλέπουμε τα αποτελέσματα του ερωτήματος για τις τρεις τελευταίες δεκαετίες. Παρατηρείται μία τρομερή αύξηση στην εισροή μεταναστών κατά την τελευταία εικοσαετία. Ειδικά η διαφορά ανάμεσα στη δεκαετία του '80 και του '90 είναι παραπάνω από 1000%, ενώ ανάμεσα στη δεκαετία του 1990 και σε αυτή του 2000 υπάρχει μία αύξηση του 80%.

ETOS	ASFAL FOREAS
2008	All
	Drop Column Field
Dekaetia	Metanasths Me
1980-1989	3629
1990-1999	44371
2000-2009	79084
Grand Total	127084

Σχ. 6-20 Πλήθος μεταναστών ανά δεκαετία εγγραφής

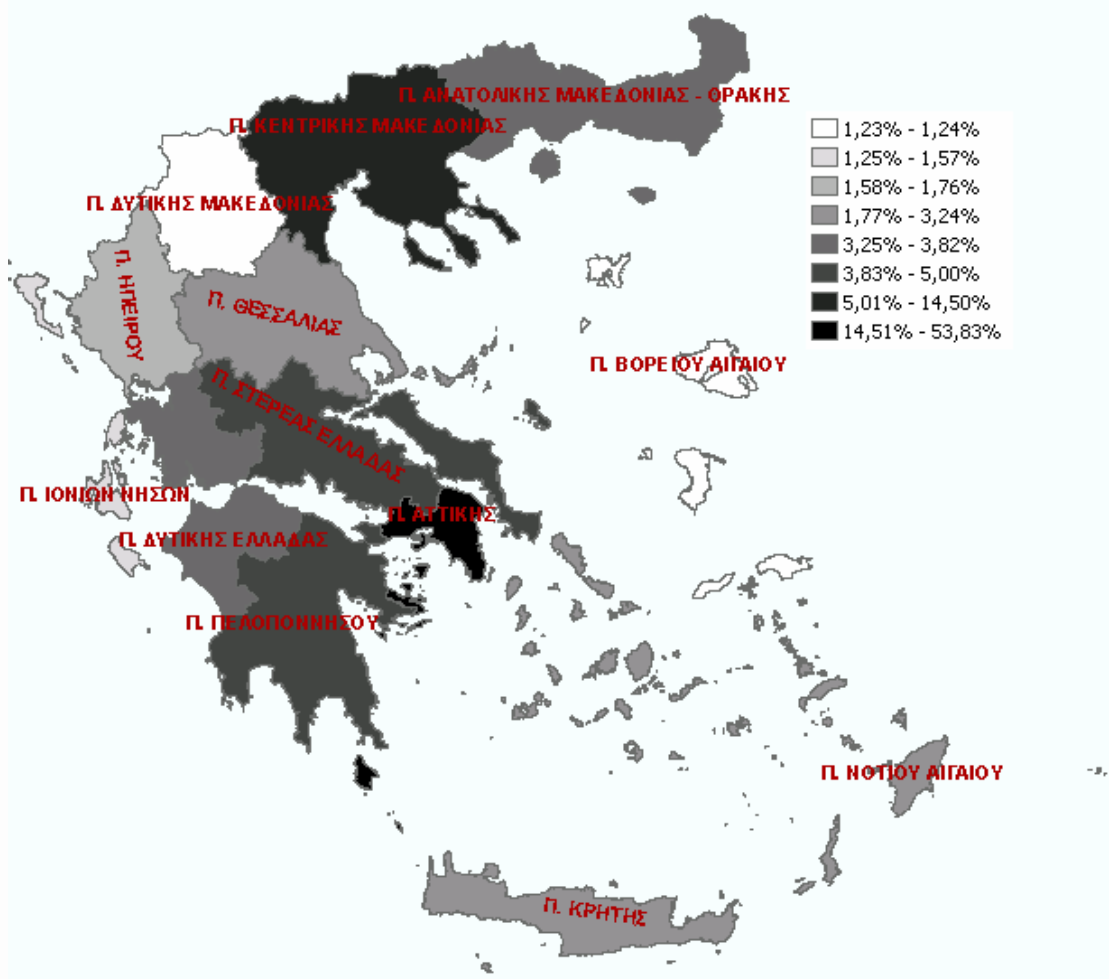
6.12 Πλήθος ατόμων ανά Έτος και ανά Περιφέρεια Εργασίας

Το τελευταίο παράδειγμα που παρουσιάζεται είναι η ποσοστωση των μεταναστών για το έτος 2008 ανά καλλικρατική περιφέρεια της Ελλάδος. Στο σχήμα 6-21 παρατηρούμε ότι την πλειονότητα αυτών τους δέχεται η περιφέρεια Αττικής (54%), ενώ ακολουθεί η περιφέρεια κεντρικής Μακεδονίας. Όλες οι υπόλοιπες περιφέρειες έχουν περίπου τον ίδιο αριθμό μεταναστών.

ASFAL FOREAS	ETOS	
All	2008	Grand Total
DIAMERISMA	Metanasths Me	Metanasths Me
ΑΝ. ΜΑΚΕΔ. ΘΡΑΚ	3,82%	3,82%
ΑΤΤΙΚΗΣ	53,83%	53,83%
Β. ΑΙΓΑΙΟΥ	1,23%	1,23%
ΔΥΤ. ΕΛΛΑΔΑΣ	3,40%	3,40%
ΔΥΤ. ΜΑΚΕΔΟΝΙΑΣ	1,24%	1,24%
ΗΠΕΙΡΟΥ	1,76%	1,76%
ΘΕΣΣΑΛΙΑΣ	2,79%	2,79%
ΙΟΝ. ΝΗΣΩΝ	1,57%	1,57%
ΚΕΝ. ΜΑΚΕΔΟΝΙΑΣ	14,50%	14,50%
ΚΡΗΤΗΣ	3,24%	3,24%
Ν. ΑΙΓΑΙΟΥ	3,07%	3,07%
ΠΕΛΟΠΟΝΝΗΣΟΥ	5,00%	5,00%
ΣΤ. ΕΛΛΑΔΑΣ	4,56%	4,56%
Grand Total	100,00%	100,00%

Σχ. 6-21 Ποσοστό μεταναστών επί του συνόλου ανά περιφέρεια

Στο σχήμα 6-22 βλέπουμε το χάρτη των περιφερειών της Ελλάδος χρωματισμένο ανάλογα με την ποσόστωση των μεταναστών.



Σχ. 6-22 Χάρτης περιφερειών Ελλάδος με ποσοστό μεταναστών επί του συνόλου

7 Βιβλιογραφικές αναφορές

1. www.oaee.gr
2. <http://geodata.gov.gr/geodata/>
3. <http://geocommons.com>
4. <http://pubs.usgs.gov>
5. Jacobson, Reed, Misner, Stacia: Microsoft SQL SERVER 2005 ANALYSIS SERVICES Step by Step, Microsoft Press.
6. Τίμος Σελλής : Προχωρημένα Θέματα βάσεων, Συμπληρωματικές σημειώσεις
7. Silberschatz , Korth , Sudarshan : Συστήματα Βάσεων δεδομένων 4^η έκδοση
8. Fon Silvers: Building and maintaining a Data warehouse
9. Ralph Kimball, Margy Ross : The Data Warehouse Toolkit Second Edition , The Complete Guide to Dimensional Modeling
10. Sybase SQL Server Transact-SQL User's Guide
11. <http://www.imepo.gr/news-competitions-gr.htm>
12. <http://technet.microsoft.com/en-us/library>
13. <http://www.spatialbi.org/>
14. Ιωσηφίδης Ελευθέριος: Διαχείριση Πολυδιάστατων Δεδομένων: Πειραματική και Συγκριτική Αξιολόγηση της Απόδοσης Εμπορικών και Ανοικτού Κώδικα DBMS (2010)
15. Τ.Δασκαλάκης, Α.Ζύσου , Σ.Καλογεράκης, Β.Κοτμανατζόγλου, Σ.Λένη, Α.Τσατσαρής: Δημιουργία, Επεξεργασία και ανάλυση περιβαλλοντικών, πολεοδομικών και χωροταξικών δεδομένων με τη χρήση γεωγραφικών συστημάτων πληροφοριών (G.I.S.)

8 ΠΑΡΑΡΤΗΜΑ

Δημιουργία Πινάκων βάσης δεδομένων

```
CREATE TABLE [dbo].[ypokatasthma](
    [ypokatasthma_id] [int] NOT NULL,
    [ypokatasthma_descr] [nvarchar](60) COLLATE Greek_CI_AS NULL,
    [di euthinsh] [nvarchar](25) COLLATE Greek_CI_AS NULL,
    [arithmos] [nvarchar](10) COLLATE Greek_CI_AS NULL,
    [tk_ypokatasthmatos] [nvarchar](5) COLLATE Greek_CI_AS NULL CONSTRAINT
[DF_ypokatasthma_tk_ypokatasthmatos] DEFAULT ((0)),
    CONSTRAINT [PK_ypokatasthma] PRIMARY KEY CLUSTERED
(
    [ypokatasthma_id] ASC
)WITH (IGNORE_DUP_KEY = OFF) ON [PRIMARY]
) ON [PRIMARY]
```

```
ALTER TABLE [dbo].[ypokatasthma] WITH CHECK ADD CONSTRAINT
[FK_ypokatasthma_tax_kwdikes] FOREIGN KEY([tk_ypokatasthmatos])
REFERENCES [dbo].[tax_kwdikes] ([TK])
```

```
CREATE TABLE [dbo].[allos_foreas](
    [foreas_id] [int] NOT NULL,
    [foreas_desc] [nvarchar](10) COLLATE Greek_CI_AS NULL,
    [xwra_id] [nvarchar](2) COLLATE Greek_CI_AS NOT NULL,
    CONSTRAINT [PK_allos_foreas] PRIMARY KEY CLUSTERED
(
    [foreas_id] ASC
)WITH (IGNORE_DUP_KEY = OFF) ON [PRIMARY]
) ON [PRIMARY]
```

```
ALTER TABLE [dbo].[allos_foreas] WITH CHECK ADD CONSTRAINT [FK_allos_foreas_xwra]
FOREIGN KEY([xwra_id])
REFERENCES [dbo].[xwra] ([xwra_id])
```

```
CREATE TABLE [dbo].[asfalism_foreas](
    [ama] [nvarchar](9) COLLATE Greek_CI_AS NOT NULL,
    [foreas_id] [int] NOT NULL,
    [flag_en] [nvarchar](1) COLLATE Greek_CI_AS NULL,
    CONSTRAINT [PK_asfalism_foreas] PRIMARY KEY CLUSTERED
(
    [ama] ASC,
    [foreas_id] ASC
)WITH (IGNORE_DUP_KEY = OFF) ON [PRIMARY]
) ON [PRIMARY]
```

```
ALTER TABLE [dbo].[asfalism_foreas] WITH CHECK ADD CONSTRAINT
[FK_asfalism_foreas_allos_foreas] FOREIGN KEY([foreas_id])
REFERENCES [dbo].[allos_foreas] ([foreas_id])
ALTER TABLE [dbo].[asfalism_foreas] WITH CHECK ADD CONSTRAINT
[FK_asfalism_foreas_Asfalismenos] FOREIGN KEY([ama])
REFERENCES [dbo].[Asfalismenos] ([AMA])
```

```

CREATE TABLE [dbo].[Asfalismenos](
    [AMA] [nchar](9) COLLATE Greek_CI_AS NOT NULL,
    [Simerino_Eponymo] [nchar](50) COLLATE Greek_CI_AS NULL,
    [Kyrio_Onoma] [nchar](30) COLLATE Greek_CI_AS NULL,
    [Eponymo_Patros] [nchar](50) COLLATE Greek_CI_AS NULL,
    [Eponymo_Mitros] [nchar](50) COLLATE Greek_CI_AS NULL,
    [Eponymo_Sizigou] [nchar](50) COLLATE Greek_CI_AS NULL,
    [Onoma_Sigigou] [nchar](30) COLLATE Greek_CI_AS NULL,
    [Etos_Gennisis] [int] NOT NULL,
    [xwra_id] [nchar](2) COLLATE Greek_CI_AS NOT NULL,
    [Filo_Asfalismenou] [nchar](1) COLLATE Greek_CI_AS NOT NULL,
    [Odos_katoikias] [nchar](50) COLLATE Greek_CI_AS NULL,
    [Arithmos_katoikias] [nchar](5) COLLATE Greek_CI_AS NULL,
    [TK_katoikias] [nchar](5) COLLATE Greek_CI_AS NOT NULL,
    [Arithmos_tautotitas] [nchar](20) COLLATE Greek_CI_AS NULL,
    [AFM] [nchar](12) COLLATE Greek_CI_AS NULL,
    [Armodia_DOY] [nchar](20) COLLATE Greek_CI_AS NULL,
    [Oikogeneiakh_katastash] [varchar](1) COLLATE Greek_CI_AS NULL,
    [Prostatevomena_melh] [varchar](2) COLLATE Greek_CI_AS NULL,
    [Odos_ergasias] [nchar](10) COLLATE Greek_CI_AS NULL,
    [Arithmos_ergasias] [nchar](10) COLLATE Greek_CI_AS NULL,
    [Tk_ergasias] [nchar](5) COLLATE Greek_CI_AS NOT NULL,
    [Epaggelma] [nchar](50) COLLATE Greek_CI_AS NULL,
    [Hmeromhnia_enarkshs_afalishs] [datetime] NULL,
    CONSTRAINT [PK_Asfalismenos] PRIMARY KEY CLUSTERED
(
    [AMA] ASC
)WITH (IGNORE_DUP_KEY = OFF) ON [PRIMARY]
) ON [PRIMARY]

ALTER TABLE [dbo].[Asfalismenos] WITH CHECK ADD CONSTRAINT
[FK_Asfalismenos_tax_kwdikes] FOREIGN KEY([TK_katoikias])
REFERENCES [dbo].[tax_kwdikes] ([TK])
ALTER TABLE [dbo].[Asfalismenos] WITH CHECK ADD CONSTRAINT
[FK_Asfalismenos_taxykwd_ypokatasthma] FOREIGN KEY([Tk_ergasias])
REFERENCES [dbo].[taxykwd_ypokatasthma] ([tk_ergasias])
ALTER TABLE [dbo].[Asfalismenos] WITH CHECK ADD CONSTRAINT
[FK_Asfalismenos_xwra] FOREIGN KEY([xwra_id])
REFERENCES [dbo].[xwra] ([xwra_id])

CREATE TABLE [dbo].[Kathgoria_klashs](
    [Oikonomikh_klash_kat] [nchar](10) COLLATE Greek_CI_AS NOT NULL,
    [Oikonomikh_klash_kat_descr] [nchar](100) COLLATE Greek_CI_AS NULL,
    CONSTRAINT [PK_Kathgoria_klashs] PRIMARY KEY CLUSTERED
(
    [Oikonomikh_klash_kat] ASC
)WITH (IGNORE_DUP_KEY = OFF) ON [PRIMARY]
) ON [PRIMARY]

CREATE TABLE [dbo].[nomoi](
    [nomos_id] [int] NOT NULL,
    [nomos_descr] [nchar](20) COLLATE Greek_CI_AS NULL,
    [diamerisma] [nchar](15) COLLATE Greek_CI_AS NULL,
    CONSTRAINT [PK_nomoi] PRIMARY KEY CLUSTERED
(

```

```

        [nomos_id] ASC
    )WITH (IGNORE_DUP_KEY = OFF) ON [PRIMARY]
    ) ON [PRIMARY]

```

```

CREATE TABLE [dbo].[Oikonomika_klashs](
    [Oikonomikh_Klash_id] [nchar](10) COLLATE Greek_CI_AS NOT NULL,
    [Etos] [nchar](10) COLLATE Greek_CI_AS NOT NULL,
    [Poso] [decimal](18, 2) NULL,
    CONSTRAINT [PK_Oikonomika_klashs] PRIMARY KEY CLUSTERED
    (
        [Oikonomikh_Klash_id] ASC,
        [Etos] ASC
    )WITH (IGNORE_DUP_KEY = OFF) ON [PRIMARY]
    ) ON [PRIMARY]

```

```

ALTER TABLE [dbo].[Oikonomika_klashs] WITH CHECK ADD CONSTRAINT
[FK_Oikonomika_klashs_Oikonomikh_klash] FOREIGN KEY([Oikonomikh_Klash_id])
REFERENCES [dbo].[Oikonomikh_klash] ([Oikonomikh_klash_id])

```

```

CREATE TABLE [dbo].[Oikonomikh_drast](
    [oikonomikh_drastiriothta_id] [nchar](10) COLLATE Greek_CI_AS NOT NULL,
    [perigrafh_drastiriohtas] [nchar](20) COLLATE Greek_CI_AS NULL,
    [kathgoria_drasthriohtas] [nchar](10) COLLATE Greek_CI_AS NULL,
    CONSTRAINT [PK_OIKONOMIKH_DRAST] PRIMARY KEY CLUSTERED
    (
        [oikonomikh_drastiriothta_id] ASC
    )WITH (IGNORE_DUP_KEY = OFF) ON [PRIMARY]
    ) ON [PRIMARY]

```

```

CREATE TABLE [dbo].[Oikonomikh_klash](
    [Oikonomikh_klash_id] [nchar](10) COLLATE Greek_CI_AS NOT NULL,
    [Oikonomikh_klash_kat] [nchar](10) COLLATE Greek_CI_AS NULL,
    CONSTRAINT [PK_Oikonomikh_klash] PRIMARY KEY CLUSTERED
    (
        [Oikonomikh_klash_id] ASC
    )WITH (IGNORE_DUP_KEY = OFF) ON [PRIMARY]
    ) ON [PRIMARY]

```

```

ALTER TABLE [dbo].[Oikonomikh_klash] WITH CHECK ADD CONSTRAINT
[FK_Oikonomikh_klash_Kathgoria_klashs] FOREIGN KEY([Oikonomikh_klash_kat])
REFERENCES [dbo].[Kathgoria_klashs] ([Oikonomikh_klash_kat])

```

```

CREATE TABLE [dbo].[Pinakas_Metavolwn](
    [AMA] [nchar](9) COLLATE Greek_CI_AS NOT NULL,
    [Etos] [int] NOT NULL,
    [Minas] [int] NOT NULL,
    [Oikonomikh_klash_id] [nchar](10) COLLATE Greek_CI_AS NOT NULL,
    CONSTRAINT [PK_Pinakas_Metavolwn] PRIMARY KEY CLUSTERED
    (
        [AMA] ASC,
        [Etos] ASC,
        [Minas] ASC
    )WITH (IGNORE_DUP_KEY = OFF) ON [PRIMARY]
    ) ON [PRIMARY]

```

```
ALTER TABLE [dbo].[Pinakas_Metavolwn] WITH CHECK ADD CONSTRAINT
[FK_Pinakas_Metavolwn_Asfalismenos] FOREIGN KEY([AMA])
REFERENCES [dbo].[Asfalismenos] ([AMA])
ALTER TABLE [dbo].[Pinakas_Metavolwn] WITH CHECK ADD CONSTRAINT
[FK_Pinakas_Metavolwn_Oikonomikh_klash] FOREIGN KEY([Oikonomikh_klash_id])
REFERENCES [dbo].[Oikonomikh_klash] ([Oikonomikh_klash_id])
```

```
CREATE TABLE [dbo].[tax_kwdikes](
    [TK] [nchar](5) COLLATE Greek_CI_AS NOT NULL,
    [nomos_id] [int] NULL,
    CONSTRAINT [PK_katoikia] PRIMARY KEY CLUSTERED
(
    [TK] ASC
)WITH (IGNORE_DUP_KEY = OFF) ON [PRIMARY]
) ON [PRIMARY]
```

```
ALTER TABLE [dbo].[tax_kwdikes] WITH CHECK ADD CONSTRAINT
[FK_tax_kwdikes_nomoi] FOREIGN KEY([nomos_id])
REFERENCES [dbo].[nomoi] ([nomos_id])
```

```
CREATE TABLE [dbo].[taxykw_dypokatasthma](
    [tk_ergasias] [nchar](5) COLLATE Greek_CI_AS NOT NULL,
    [ypokatasthma_id] [int] NOT NULL,
    CONSTRAINT [PK_taxykw_dypokatasthma_1] PRIMARY KEY CLUSTERED
(
    [tk_ergasias] ASC
)WITH (IGNORE_DUP_KEY = OFF) ON [PRIMARY]
) ON [PRIMARY]
```

```
ALTER TABLE [dbo].[taxykw_dypokatasthma] WITH CHECK ADD CONSTRAINT
[FK_taxykw_dypokatasthma_tax_kwdikes] FOREIGN KEY([tk_ergasias])
REFERENCES [dbo].[tax_kwdikes] ([TK])
ALTER TABLE [dbo].[taxykw_dypokatasthma] WITH CHECK ADD CONSTRAINT
[FK_taxykw_dypokatasthma_ypokatasthma] FOREIGN KEY([ypokatasthma_id])
REFERENCES [dbo].[ypokatasthma] ([ypokatasthma_id])
```

```
CREATE TABLE [dbo].[Xwra_Organismos](
    [Xwra_id] [nchar](2) COLLATE Greek_CI_AS NOT NULL,
    [Organismos] [nchar](2) COLLATE Greek_CI_AS NOT NULL,
    CONSTRAINT [PK_Xwra_Organismos] PRIMARY KEY CLUSTERED
(
    [Xwra_id] ASC,
    [Organismos] ASC
)WITH (IGNORE_DUP_KEY = OFF) ON [PRIMARY]
) ON [PRIMARY]
```

```
ALTER TABLE [dbo].[Xwra_Organismos] WITH CHECK ADD CONSTRAINT
[FK_Xwra_Organismos_xwra1] FOREIGN KEY([Xwra_id])
REFERENCES [dbo].[xwra] ([xwra_id])
```

```
CREATE TABLE [dbo].[xwra](
    [xwra_id] [nchar](2) COLLATE Greek_CI_AS NOT NULL,
    [xwra_eng] [nchar](52) COLLATE Greek_CI_AS NULL,
```

```

        [xwra_el] [nchar](70) COLLATE Greek_CI_AS NULL,
        [hpeiros] [nchar](2) COLLATE Greek_CI_AS NULL,
    CONSTRAINT [PK_xwra] PRIMARY KEY CLUSTERED
    (
        [xwra_id] ASC
    )WITH (IGNORE_DUP_KEY = OFF) ON [PRIMARY]
    ) ON [PRIMARY]

```

Δημιουργία Πινάκων Αποθήκης δεδομένων

```

CREATE TABLE [dbo].[DIM_GEOGRAPHY](
    [NOMOS_ID] [int] NOT NULL,
    [NOMOS_DESC] [nchar](20) COLLATE Greek_CI_AS NULL,
    [DIAMERISMA] [nchar](15) COLLATE Greek_CI_AS NULL,
    CONSTRAINT [PK_DIM_GEOGRAPHY] PRIMARY KEY CLUSTERED
    (
        [NOMOS_ID] ASC
    )WITH (IGNORE_DUP_KEY = OFF) ON [PRIMARY]
    ) ON [PRIMARY]

```

```

CREATE TABLE [dbo].[DIM_METANASTHS](
    [KWD_METANASTH] [nchar](9) COLLATE Greek_CI_AS NOT NULL,
    [ETOS_GENHSHS] [int] NULL,
    [KWD_XWRA_KATAG] [nchar](2) COLLATE Greek_CI_AS NULL,
    [FYLO] [nchar](1) COLLATE Greek_CI_AS NULL,
    [ETOS_APOGRAFHS] [int] NULL,
    [KWD_OIKOG_KATAST] [nchar](10) COLLATE Greek_CI_AS NULL,
    CONSTRAINT [PK_DIM_METANASTHS] PRIMARY KEY CLUSTERED
    (
        [KWD_METANASTH] ASC
    )WITH (IGNORE_DUP_KEY = OFF) ON [PRIMARY]
    ) ON [PRIMARY]

```

```

ALTER TABLE [dbo].[DIM_METANASTHS] WITH CHECK ADD CONSTRAINT
[FK_DIM_METANASTHS_DIM_OIKOGEN_KATAST] FOREIGN
KEY([KWD_OIKOG_KATAST])
REFERENCES [dbo].[DIM_OIKOGEN_KATAST] ([KWD_OIKOG_KATAST])
ALTER TABLE [dbo].[DIM_METANASTHS] WITH CHECK ADD CONSTRAINT
[FK_DIM_METANASTHS_DIM_XWRA_KAT] FOREIGN KEY([KWD_XWRA_KATAG])
REFERENCES [dbo].[DIM_XWRA_KAT] ([KWD_XWRA_KATAG])

```

```

CREATE TABLE [dbo].[DIM_OIKOGEN_KATAST](
    [KWD_OIKOG_KATAST] [nchar](10) COLLATE Greek_CI_AS NOT NULL,
    [DESC_OIKOG_KATAST] [nchar](15) COLLATE Greek_CI_AS NULL,
    CONSTRAINT [PK_DIM_OIKOGEN_KATAST] PRIMARY KEY CLUSTERED
    (
        [KWD_OIKOG_KATAST] ASC
    )WITH (IGNORE_DUP_KEY = OFF) ON [PRIMARY]
    ) ON [PRIMARY]
CREATE TABLE [dbo].[DIM_OIKONOM_DRAST](
    [KWD_OIKONOM_DRAST] [int] NOT NULL,
    [DESC_OIKONOM_DRAST] [nchar](150) COLLATE Greek_CI_AS NULL,
    CONSTRAINT [PK_DIM_OIKONOM_DRAST] PRIMARY KEY CLUSTERED
    (
        [KWD_OIKONOM_DRAST] ASC
    )WITH (IGNORE_DUP_KEY = OFF) ON [PRIMARY]

```



```
) ON [PRIMARY]
```

```
CREATE TABLE [dbo].[DIM_TIME](
    [KWD_HMEROM] [int] IDENTITY(1,1) NOT NULL,
    [EXAMHNO] [int] NULL,
    [ETOS] [int] NULL,
    CONSTRAINT [PK_DIM_TIME] PRIMARY KEY CLUSTERED
    (
        [KWD_HMEROM] ASC
    )WITH (IGNORE_DUP_KEY = OFF) ON [PRIMARY]
) ON [PRIMARY]
```

```
CREATE TABLE [dbo].[DIM_XWRA_KAT](
    [KWD_XWRA_KATAG] [nchar](2) COLLATE Greek_CI_AS NOT NULL,
    [DESC_XWRA_KATAG] [nchar](70) COLLATE Greek_CI_AS NULL,
    [DESC_XWRA_KATAG_ENG] [nchar](52) COLLATE Greek_CI_AS NULL,
    [ORGANISMOS] [nchar](10) COLLATE Greek_CI_AS NULL,
    [HPEIROS] [nchar](2) COLLATE Greek_CI_AS NULL,
    CONSTRAINT [PK_DIM_XWRA_KAT] PRIMARY KEY CLUSTERED
    (
        [KWD_XWRA_KATAG] ASC
    )WITH (IGNORE_DUP_KEY = OFF) ON [PRIMARY]
) ON [PRIMARY]
```

```
CREATE TABLE [dbo].[DIM_YPOKATAST_FOREAS](
    [KWD_YPOKATASTHMATOS] [int] NOT NULL,
    [DESC_YPOKATASTHMATOS] [nchar](60) COLLATE Greek_CI_AS NULL,
    [ASFAL_FOREAS] [nchar](10) COLLATE Greek_CI_AS NULL,
    CONSTRAINT [PK_DIM_YPOKATAST_FOREAS] PRIMARY KEY CLUSTERED
    (
        [KWD_YPOKATASTHMATOS] ASC
    )WITH (IGNORE_DUP_KEY = OFF) ON [PRIMARY]
) ON [PRIMARY]
```

```
CREATE TABLE [dbo].[FACT_PERIODIKH_KINHSH](
    [KWD_METANASTH] [nchar](9) COLLATE Greek_CI_AS NOT NULL,
    [KWD_PER_KATOIKIAS] [int] NULL,
    [KWD_PER_ERGASIAS] [int] NULL,
    [KWD_YPOKATASTHMATOS] [int] NULL,
    [KWD_OIKONOM_DRAST] [int] NULL,
    [KWD_HMEROM] [int] NOT NULL,
    [HMERES_ASFALISHS] [nchar](10) COLLATE Greek_CI_AS NULL,
    [APODOXES] [money] NULL,
    [EISFORES] [money] NULL,
    CONSTRAINT [PK_FACT_PERIODIKH_KINHSH] PRIMARY KEY CLUSTERED
    (
        [KWD_METANASTH] ASC,
        [KWD_HMEROM] ASC
    )WITH (IGNORE_DUP_KEY = OFF) ON [PRIMARY]
) ON [PRIMARY]
```

```
ALTER TABLE [dbo].[FACT_PERIODIKH_KINHSH] WITH CHECK ADD CONSTRAINT
[FK_FACT_PERIODIKH_KINHSH_DIM_GEOGRAPHY] FOREIGN
KEY([KWD_PER_KATOIKIAS])
```

```

REFERENCES [dbo].[DIM_GEOGRAPHY] ([NOMOS_ID])
ALTER TABLE [dbo].[FACT_PERIODIKH_KINHSH] WITH CHECK ADD CONSTRAINT
[FK_FACT_PERIODIKH_KINHSH_DIM_GEOGRAPHY1] FOREIGN
KEY([KWD_PER_ERGASIAS])
REFERENCES [dbo].[DIM_GEOGRAPHY] ([NOMOS_ID])
ALTER TABLE [dbo].[FACT_PERIODIKH_KINHSH] WITH CHECK ADD CONSTRAINT
[FK_FACT_PERIODIKH_KINHSH_DIM_METANASTHS] FOREIGN
KEY([KWD_METANASTH])
REFERENCES [dbo].[DIM_METANASTHS] ([KWD_METANASTH])
ALTER TABLE [dbo].[FACT_PERIODIKH_KINHSH] WITH CHECK ADD CONSTRAINT
[FK_FACT_PERIODIKH_KINHSH_DIM_OIKONOM_DRAST] FOREIGN
KEY([KWD_OIKONOM_DRAST])
REFERENCES [dbo].[DIM_OIKONOM_DRAST] ([KWD_OIKONOM_DRAST])
ALTER TABLE [dbo].[FACT_PERIODIKH_KINHSH] WITH CHECK ADD CONSTRAINT
[FK_FACT_PERIODIKH_KINHSH_DIM_TIME] FOREIGN KEY([KWD_HMEROM])
REFERENCES [dbo].[DIM_TIME] ([KWD_HMEROM])
ALTER TABLE [dbo].[FACT_PERIODIKH_KINHSH] WITH CHECK ADD CONSTRAINT
[FK_FACT_PERIODIKH_KINHSH_DIM_YPOKATAST_FOREAS] FOREIGN
KEY([KWD_YPOKATASTHMATOS])
REFERENCES [dbo].[DIM_YPOKATAST_FOREAS] ([KWD_YPOKATASTHMATOS])

```

Scripts εισαγωγής δεδομένων πίνακα μεταβολών σε Βάση Δεδομένων

```

DECLARE @ama int, @klash_id nvarchar(5), @hmdi datetime, @hmep datetime, @enasf
datetime
DECLARE @etos_apo int, @etos_ews int, @mhnas_apo int, @mhnas_ews int
, @counter_mhnas int
DECLARE asfacursor Cursor FOR select asfaamas, asfaaska as
aska, convert(datetime, asfahmdi, 103) as hmdi, convert(datetime, asfahmep, 103) as
hmep, ASFAENASF as enasf from temp_tebe.dbo.asfalismenoi
where asfahmdi is not null and asfahmep is not null order by asfaamas
OPEN ASFACURSOR

FETCH NEXT FROM asfacursor into @ama, @klash_id, @hmdi, @hmep, @enasf

While (@@FETCH_STATUS = 0)
Begin

    IF @hmdi is not null and @hmep is not null
    BEGIN
        IF datepart(yyyy, @enasf) < 2007 and datepart(yyyy, @hmdi) < 2007 and
datepart(yyyy, @hmep) < 2007 --1
        BEGIN
            SET @etos_apo = 2007
            SET @mhnas_apo = 1
            SET @etos_ews = 2009
            SET @mhnas_ews = 10
        -----
        IF @etos_ews >= 2007
        BEGIN
            SET @counter_mhnas = 0
            WHILE @etos_apo <= @etos_ews
            BEGIN
                IF @counter_mhnas > 12
                    SET @counter_mhnas = 1
                else
                    SET @counter_mhnas = @mhnas_apo
            
```

```

        WHILE (@counter_mhnas<13)
        BEGIN
            INSERT INTO tebe.dbo.pinakas_metavolwn values
            (@ama,@etos_apo,@counter_mhnas,@klash_id)
            --PRINT
            convert(nchar(8),@ama)+convert(nchar(5),@etos_apo)+convert(nchar(3),@counter_mhnas)+c
            onvert(nchar(5),@etos_ews)+convert(nchar(3),@mhnas_ews)+convert(nchar(11),@enasf,103)
            +isnull(convert(nvarchar(10),@hmdi,103),' ')

            IF @etos_apo=@etos_ews and
            @counter_mhnas=@mhnas_ews
            BREAK
            SET @counter_mhnas=@counter_mhnas+1
        END
        SET @etos_apo=@etos_apo+1
    END
END
-----
END
IF datepart(yyyy,@enasf)<2007 and datepart(yyyy,@hmdi)<2007 and
datepart(yyyy,@hmep)>=2007 --2
BEGIN
    SET @etos_apo=datepart(yyyy,@hmep)
    SET @mhnas_apo=datepart(mm,@hmep)
    SET @etos_ews=2009
    SET @mhnas_ews=10
-----
    IF @etos_ews>=2007
    BEGIN
        SET @counter_mhnas=0
        WHILE @etos_apo<=@etos_ews
        BEGIN
            IF @counter_mhnas>12
                SET @counter_mhnas=1
            else
                SET @counter_mhnas=@mhnas_apo
            WHILE (@counter_mhnas<13)
            BEGIN
                INSERT INTO tebe.dbo.pinakas_metavolwn values
                (@ama,@etos_apo,@counter_mhnas,@klash_id)
                --PRINT
                convert(nchar(8),@ama)+convert(nchar(5),@etos_apo)+convert(nchar(3),@counter_mhnas)+c
                onvert(nchar(5),@etos_ews)+convert(nchar(3),@mhnas_ews)+convert(nchar(11),@enasf,103)
                +isnull(convert(nvarchar(10),@hmdi,103),' ')

                IF @etos_apo=@etos_ews and
                @counter_mhnas=@mhnas_ews
                BREAK
                SET @counter_mhnas=@counter_mhnas+1
            END
            SET @etos_apo=@etos_apo+1
        END
    END
END
-----
END
IF datepart(yyyy,@enasf)<2007 and datepart(yyyy,@hmdi)>=2007 --3
BEGIN
    SET @etos_apo=2007

```

```

SET @mhnas_apo=1
SET @etos_ews=datepart(yyyy,@hmdi)
SET @mhnas_ews=datepart(mm,@hmdi)
-----
IF @etos_ews>=2007
BEGIN
    SET @counter_mhnas=0
    WHILE @etos_apo<=@etos_ews
    BEGIN
        IF @counter_mhnas>12
            SET @counter_mhnas=1
        else
            SET @counter_mhnas=@mhnas_apo
        WHILE (@counter_mhnas<13)
        BEGIN
            INSERT INTO tebe.dbo.pinakas_metavolwn values
            (@ama,@etos_apo,@counter_mhnas,@klash_id)
            --PRINT
            convert(nchar(8),@ama)+convert(nchar(5),@etos_apo)+convert(nchar(3),@counter_mhnas)+c
            onvert(nchar(5),@etos_ews)+convert(nchar(3),@mhnas_ews)+convert(nchar(11),@enasf,103)
            +isnull(convert(nvarchar(10),@hmdi,103),' ')

            IF @etos_apo=@etos_ews and
            @counter_mhnas=@mhnas_ews
                BREAK
            SET @counter_mhnas=@counter_mhnas+1
        END
        SET @etos_apo=@etos_apo+1
    END
END
-----
SET @etos_apo=datepart(yyyy,@hmep)
SET @mhnas_apo=datepart(mm,@hmep)
SET @etos_ews=2009
SET @mhnas_ews=10
-----
IF @etos_ews>=2007
BEGIN
    SET @counter_mhnas=0
    WHILE @etos_apo<=@etos_ews
    BEGIN
        IF @counter_mhnas>12
            SET @counter_mhnas=1
        else
            SET @counter_mhnas=@mhnas_apo
        WHILE (@counter_mhnas<13)
        BEGIN
            INSERT INTO tebe.dbo.pinakas_metavolwn values
            (@ama,@etos_apo,@counter_mhnas,@klash_id)
            --PRINT
            convert(nchar(8),@ama)+convert(nchar(5),@etos_apo)+convert(nchar(3),@counter_mhnas)+c
            onvert(nchar(5),@etos_ews)+convert(nchar(3),@mhnas_ews)+convert(nchar(11),@enasf,103)
            +isnull(convert(nvarchar(10),@hmdi,103),' ')

            IF @etos_apo=@etos_ews and
            @counter_mhnas=@mhnas_ews
                BREAK
            SET @counter_mhnas=@counter_mhnas+1

```

```

        END
        SET @etos_apo=@etos_apo+1
    END
END
-----
    END
    IF datepart(yyyy,@enasf)>2007 --4
    BEGIN
        SET @etos_apo=datepart(yyyy,@enasf)
        SET @mhnas_apo=datepart(mm,@enasf)
        SET @etos_ews=datepart(yyyy,@hmdi)
        SET @mhnas_ews=datepart(mm,@hmdi)
    -----
    IF @etos_ews>=2007
    BEGIN
        SET @counter_mhnas=0
        WHILE @etos_apo<=@etos_ews
        BEGIN
            IF @counter_mhnas>12
                SET @counter_mhnas=1
            else
                SET @counter_mhnas=@mhnas_apo
            WHILE (@counter_mhnas<13)
            BEGIN
                INSERT INTO tebe.dbo.pinakas_metavolwn values
                (@ama,@etos_apo,@counter_mhnas,@klash_id)
                --PRINT
                convert(nchar(8),@ama)+convert(nchar(5),@etos_apo)+convert(nchar(3),@counter_mhnas)+
                onvert(nchar(5),@etos_ews)+convert(nchar(3),@mhnas_ews)+convert(nchar(11),@enasf,103)
                +isnull(convert(nvarchar(10),@hmdi,103),'')
                IF @etos_apo=@etos_ews and
                @counter_mhnas=@mhnas_ews
                    BREAK
                SET @counter_mhnas=@counter_mhnas+1
            END
            SET @etos_apo=@etos_apo+1
        END
    END
END
-----
    SET @etos_apo=datepart(yyyy,@hmep)
    SET @mhnas_apo=datepart(mm,@hmep)
    SET @etos_ews=2009
    SET @mhnas_ews=10
    -----
    IF @etos_ews>=2007
    BEGIN
        SET @counter_mhnas=0
        WHILE @etos_apo<=@etos_ews
        BEGIN
            IF @counter_mhnas>12
                SET @counter_mhnas=1
            else
                SET @counter_mhnas=@mhnas_apo
            WHILE (@counter_mhnas<13)
            BEGIN
                INSERT INTO tebe.dbo.pinakas_metavolwn values
                (@ama,@etos_apo,@counter_mhnas,@klash_id)

```

```

--PRINT
convert(nchar(8),@ama)+convert(nchar(5),@etos_apo)+convert(nchar(3),@counter_mhnas)+
onvert(nchar(5),@etos_ews)+convert(nchar(3),@mhnas_ews)+convert(nchar(11),@enasf,103)
+isnull(convert(nvarchar(10),@hmdi,103),' ')

IF @etos_apo=@etos_ews and
@counter_mhnas=@mhnas_ews
BREAK
SET @counter_mhnas=@counter_mhnas+1
END
SET @etos_apo=@etos_apo+1
END
END
END
-----
END
END

FETCH NEXT FROM asfacursor into @ama,@klash_id,@hmdi,@hmep,@enasf

End

Close asfacursor
Deallocate asfacursor

```

ETL διαδικασία για εισαγωγή δεδομένων σε Αποθήκη Δεδομένων

```

INSERT INTO [IKA_TEBE_DW].[DBO].[FACT_PERIODIKH_KINHSH]
([KWD_METANASTH]
,[KWD_PER_KATOIKIAS]
,[KWD_PER_ERGASIAS]
,[KWD_YPOKATASTHMATOS]
,[KWD_OIKONOM_DRAST]
,[KWD_HMEROM]
,[PALAIOTHTA]
,[HMERES_ASFALISHS]
,[APODOXES]
,[EISFORES])
SELECT
ATOMIKA.AMA,KATOIKIA,ERGASIA,YPOKATASTHMA,OIKONOMIKH_DRASTHRIOTHTA,K
WD_HMEROM,PALAIOTHTA,NULL,NULL,SYNOLO FROM
(
SELECT AMA ,NOMOI_KATOIKIAS.NOMOS_IKA AS
KATOIKIA,NOMOI_ERGASIAS.NOMOS_IKA AS ERGASIA,D.YPOKATASTHMA_ID AS
YPOKATASTHMA,NULL AS OIKONOMIKH_DRASTHRIOTHTA FROM
TEBE.DBO.ASFALISMENOS A
,(SELECT C.TK,B.NOMOS_ID AS NOMOS_IKA,A.NOMOS_ID AS NOMOS_TEBE FROM
TEBE.DBO.TAX_KWDIKES C,TEBE.DBO.NOMOI A ,IKA.DBO.NOMOI B WHERE
A.NOMOS_DESCR=B.DESCRPTION AND C.NOMOS_ID=A.NOMOS_ID
UNION ALL SELECT C.TK,'310','5' FROM TEBE.DBO.NOMOI A,TEBE.DBO.TAX_KWDIKES C
WHERE A.NOMOS_ID=C.NOMOS_ID AND A.NOMOS_ID =5') NOMOI_KATOIKIAS
,(SELECT C.TK,B.NOMOS_ID AS NOMOS_IKA,A.NOMOS_ID AS NOMOS_TEBE FROM
TEBE.DBO.TAX_KWDIKES C,TEBE.DBO.NOMOI A ,IKA.DBO.NOMOI B WHERE
A.NOMOS_DESCR=B.DESCRPTION AND C.NOMOS_ID=A.NOMOS_ID
UNION ALL SELECT C.TK,'310','5' FROM TEBE.DBO.NOMOI A,TEBE.DBO.TAX_KWDIKES C

```

```

WHERE A.NOMOS_ID=C.NOMOS_ID AND A.NOMOS_ID ='5') NOMOI_ERGASIAS
,TEBE.DBO.TAXYKWD_YPOKATASTHMA D
WHERE A.TK_KATOIKIAS= NOMOI_KATOIKIAS.TK AND
A.TK_ERGASIAS=NOMOI_ERGASIAS.TK AND A.TK_ERGASIAS=D.TK_ERGASIAS
) ATOMIKA ,
(SELECT A.AMA,D.KWD_HMEROM,
CASE WHEN EXAMHNO='1' THEN (((D.ETOS*12)+6) -
((SUBSTRING(CONVERT(CHAR,HMEROMHNNIA_ENARKSHS_AFASLISHS,103),7,4)*12)+
SUBSTRING(CONVERT(CHAR,HMEROMHNNIA_ENARKSHS_AFASLISHS,103),4,2)))/12
WHEN EXAMHNO='2' THEN (((D.ETOS*12)+12) -
((SUBSTRING(CONVERT(CHAR,HMEROMHNNIA_ENARKSHS_AFASLISHS,103),7,4)*12)+
SUBSTRING(CONVERT(CHAR,HMEROMHNNIA_ENARKSHS_AFASLISHS,103),4,2)))/12
END AS PALAIOTHTA,
SUM(B.POSO) AS SYNOLO
FROM TEBE.DBO.PINAKAS_METAVOLWN A,TEBE.DBO.OIKONOMIKA_KLASHS B
,TEBE.DBO.ASFALISMENOS C ,IKA_TEBE_DW.DBO.DIM_TIME D
WHERE A.OIKONOMIKH_KLASH_ID=B.OIKONOMIKH_KLASH_ID AND A.ETOS=B.ETOS
AND A.AMA=C.AMA AND A.ETOS=D.ETOS
AND EXAMHNO = CASE WHEN MINAS BETWEEN 1 AND 6 THEN '1'
WHEN MINAS BETWEEN 7 AND 12 THEN '2'
END
GROUP BY A.AMA,D.KWD_HMEROM,
CASE WHEN EXAMHNO='1' THEN (((D.ETOS*12)+6) -
((SUBSTRING(CONVERT(CHAR,HMEROMHNNIA_ENARKSHS_AFASLISHS,103),7,4)*12)+
SUBSTRING(CONVERT(CHAR,HMEROMHNNIA_ENARKSHS_AFASLISHS,103),4,2)))/12
WHEN EXAMHNO='2' THEN (((D.ETOS*12)+12) -
((SUBSTRING(CONVERT(CHAR,HMEROMHNNIA_ENARKSHS_AFASLISHS,103),7,4)*12)+
SUBSTRING(CONVERT(CHAR,HMEROMHNNIA_ENARKSHS_AFASLISHS,103),4,2)))/12
END) OIKONOMIKA
WHERE ATOMIKA.AMA=OIKONOMIKA.AMA

INSERT INTO [IKA_TEBE_DW].[DBO].[DIM_METANASTHS]
([KWD_METANASTH]
,[HMER_GENNHSHS]
,[KWD_XWRA_KATAG]
,[FYLO]
,[ETOS_APOGRAFHS]
,[KWD_OIKOG_KATAST])
SELECT
AMA,SUBSTRING(CONVERT(VARCHAR(10),DATE_BIRTH,101),7,4),NATIONALITY_ID,
CASE SEX
WHEN 'Α' THEN 'Α'
ELSE 'Ά'
END , ETOS_APOGRAFIS , OIKOGENEIAKH_KATAST
FROM IKA.DBO.ASFALISMENOS
UNION ALL
SELECT
AMA,ETOS_GENNHISIS,XWRA_ID,FILO_ASFALISMENOU,CONVERT(INT,SUBSTRING(CON
VERT(VARCHAR(10),HMEROMHNNIA_ENARKSHS_AFASLISHS,101),7,4)),
OIKOGENEIAKH_KATASTASH
FROM TEBE.DBO.ASFALISMENOS

INSERT INTO IKA_TEBE_DW.DBO.DIM_GEOGRAPHY
SELECT RTRIM(NOMOS_ID),RTRIM(DESCRIPTION),RTRIM(DIAMERISMA) FROM
(SELECT A.*,B.DIAMERISMA FROM IKA.DBO.NOMOI A,TEBE.DBO.NOMOI B
WHERE A.DESCRPTION=B.NOMOS_DESCR

```



```

UNION ALL
SELECT A.*, 'ΣΤΕΡΕΑ ΕΛΛΑΔΑ' FROM IKA.DBO.NOMOI A WHERE A.DESCRPTION NOT IN
(SELECT NOMOS_DESCR FROM TEBE.DBO.NOMOI B)
UNION ALL
SELECT 310, NOMOS_DESCR, DIAMERISMA FROM TEBE.DBO.NOMOI WHERE
NOMOS_DESCR= 'ΑΤΤΙΚΗΣ') OLIKO
ORDER BY NOMOS_ID

```

```

INSERT INTO IKA_TEBE_DW.DBO.DIM_OIKOGEN_KATAST
SELECT DISTINCT ASFAOIKA,
CASE ASFAOIKA
WHEN 'A' THEN 'ΑΓΑΜΟΣ'
WHEN 'X' THEN 'ΧΗΡΟΣ'
WHEN 'E' THEN 'ΕΓΓΑΜΟΣ'
WHEN 'Δ' THEN 'ΔΙΑΖΕΥΜΕΝΟΣ'
END
FROM TEMP_TEBE.DBO.ASFALISMENOI WHERE ASFAOIKA BETWEEN 'A' AND 'Ω'

```

```

INSERT INTO IKA_TEBE_DW.DBO.DIM_OIKONOM_DRAST
SELECT * FROM IKA.DBO.OIKONOMIKI_DRASTIRIOTITA

```

```

INSERT INTO [IKA_TEBE_DW].[DBO].[DIM_XWRA_KAT]
([KWD_XWRA_KATAG]
, [DESC_XWRA_KATAG]
, [DESC_XWRA_KATAG_ENG]
, [ORGANISMOS]
, [HPEIROS])

```

```

SELECT
CASE WHEN A.XWRA_ID IS NOT NULL THEN A.XWRA_ID
ELSE NATIONALITY_ID
END,
CASE WHEN A.XWRA_ID IS NOT NULL THEN XWRA_EL
ELSE RTRIM(DESCRIPTION)
END,
CASE WHEN A.XWRA_ID IS NOT NULL THEN XWRA_ENG
ELSE RTRIM(DESCRIPTION)
END,
B.ORGANISMOS,
A.HPEIROS
FROM TEBE.DBO.XWRA A
FULL OUTER JOIN TEBE.DBO.XWRA_ORGANISMOS B
ON (A.XWRA_ID=B.XWRA_ID)
FULL OUTER JOIN IKA.DBO.NATIONALITY C
ON (A.XWRA_ID=C.NATIONALITY_ID)
INSERT INTO [IKA_TEBE_DW].[DBO].[DIM_YPOKATAST_FOREAS]
([KWD_YPOKATASTHMATOS]
, [DESC_YPOKATASTHMATOS]
, [ASFAL_FOREAS])
SELECT [YPOKATASTHMA_ID], [YPOKATASTHMA_DESCR], 'OAE' FROM
[TEBE].[DBO].[YPOKATASTHMA]
UNION ALL
SELECT '0', 'IKA', 'IKA ETAM'

```