

Αυτόματη παραγωγή αναφορών για την κατάσταση ενός
Key Performance Indicator

Η Μεταπτυχιακή Διπλωματική Εργασία

υποβάλλεται στην ορισθείσα

από τη Συνέλευση

του Τμήματος Μηχανικών Η/Υ και Πληροφορικής

Εξεταστική Επιτροπή

από την

Δήμητρα Γεωργιάδου

ως μέρος των υποχρεώσεων για την απόκτηση του

ΔΙΠΛΩΜΑΤΟΣ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
ΣΤΗ ΜΗΧΑΝΙΚΗ ΔΕΔΟΜΕΝΩΝ ΚΑΙ ΥΠΟΛΟΓΙΣΤΙΚΩΝ
ΣΥΣΤΗΜΑΤΩΝ

ΜΕ ΕΙΔΙΚΕΥΣΗ
ΣΤΑ ΠΡΟΗΓΜΕΝΑ ΥΠΟΛΟΓΙΣΤΙΚΑ ΣΥΣΤΗΜΑΤΑ

Πανεπιστήμιο Ιωαννίνων

Πολυτεχνική Σχολή

Ιωάννινα 2022

Εξεταστική επιτροπή:

- **Παναγιώτης Βασιλειάδης**, Καθηγητής, Τμήμα Μηχανικών Η/Υ και Πληροφορικής, Πανεπιστήμιο Ιωαννίνων (Επιβλέπων)
- **Απόστολος Ζάρρας**, Καθηγητής, Τμήμα Μηχανικών Η/Υ και Πληροφορικής, Πανεπιστήμιο Ιωαννίνων
- **Ευαγγελία Πιτουρά**, Καθηγήτρια, Τμήμα Μηχανικών Η/Υ και Πληροφορικής, Πανεπιστήμιο Ιωαννίνων

ΑΦΙΕΡΩΣΗ

Σε όσους δεν παύουν να προσπαθούν να γίνουν καλύτεροι.

ΕΥΧΑΡΙΣΤΙΕΣ

Ευχαριστώ όλους του κοντινούς μου ανθρώπους που σε κάθε δυσκολία που συνάντησα ήταν πάντα εκεί με ένα καλό λόγο και μου έδωσαν δύναμη να συνεχίσω.

Η ολοκλήρωση της διατριβής θα ήταν αδύνατη χωρίς την πολύτιμη βοήθεια του καθηγητή μου, Παναγιώτη Βασιλειάδη, του εκφράζω ένα βαθύ ευχαριστώ για την υποστήριξη που μου πρόσφερε.

ΠΕΡΙΕΧΟΜΕΝΑ

Αφιερωση	i
Ευχαριστιες	i
Περιεχομενα	i
Καταλογος Σχηματων	iii
Καταλογος Αλγοριθμων	v
Περίληψη	vi
Extended Abstract	vii
ΚΕΦ ΑΛΑΙΟ 1 Εισαγωγή	1
1.1 Στόχοι.....	1
1.1 Δομή Διατριβής.....	2
ΚΕΦ ΑΛΑΙΟ 2 ΘΕΩΡΗΤΙΚΟ ΥΠΟΒΑΘΡΟ	4
2.1 Data warehouses	4
2.2 Κύβος (Cube).....	5
2.3 Διαστάσεις (Dimensions)	6
2.4 Ιεραρχίες (Hierarchies)	6
2.5 Star-schema.....	7
2.6 Βασικός δείκτης απόδοσης (Key Performance Indicator).....	8
2.7 KPIs in Analysis Services.....	9
2.8 Μελέτη UX KPI με βάση το ερωτηματολόγιο εμπειρίας χρήστη	14
2.9 Μελέτη ανάλυσης επιχειρηματικών διαδικασιών με αυτόματο εντοπισμό κατωφλίων KPI	15
2.10 Ενίσχυση της λήψης αποφάσεων μέσω δια δραστικής ανάλυσης What- If.....	15

2.11	Μελέτη βελτίωσης της λήψης επιχειρηματικών αποφάσεων με βάση τους KPI από ένα σύστημα διαχείρισης.....	16
2.12	Boomerang: Προληπτικές συστάσεις με βάση τη διορατικότητα για την ανάλυση των δεδομένων συνομιλίας.....	17
2.13	Μελέτη MetaInsight: Αυτόματη ανακάλυψη δομημένης διερευνητικής γνώσης για διερευνητική ανάλυση δεδομένων	17
ΚΕΦΑΛΑΙΟ 3	Μηχανισμός Αυτόματης Παραγωγής Αναφορών	19
3.1	Παραγωγή KPI.....	19
3.2	Σύγκριση KPI με παρόμοιες τιμές.....	22
3.3	Σύγκριση της πραγματικής τιμής του KPI με την αναμενόμενη τιμή ως προς το ανώτερο επίπεδο ή ως προς το σύνολο	30
3.4	SQL Ερωτήσεις.....	37
3.5	Παραδείγματα.....	39
ΚΕΦΑΛΑΙΟ 4	Υλοποίηση	49
4.1	Πλατφόρμα και Προγραμματιστικά εργαλεία	49
4.2	Λεπτομέρειες Υλοποίησης.....	50
4.3	Αρχιτεκτονική του Λογισμικού	51
4.4	Έλεγχος αποτελεσμάτων	55
ΚΕΦΑΛΑΙΟ 5	Πειράματα	56
5.1	Σύγκριση KPI με παρόμοιες τιμές, εκτέλεση ερώτησης για κάθε παρόμοια τιμή.....	56
5.2	Σύγκριση KPI με παρόμοιες τιμές, εκτέλεση μια ερώτησης με τιμή το ανώτερο επίπεδο.....	63
5.3	Σύγκριση της πραγματικής τιμής του KPI με την αναμενόμενη τιμή ως προς το ανώτερο επίπεδο και ως προς το σύνολο	67
ΚΕΦΑΛΑΙΟ 6	Συμπεράσματα	69
6.1	Σύνοψη αποτελεσμάτων	69
6.2	Μελλοντικές επεκτάσεις.....	70
	Βιβλιογραφία	72
	Συντομο Βιογραφικό	74

ΚΑΤΑΛΟΓΟΣ ΣΧΗΜΑΤΩΝ

Σχήμα 2.1 Παράδειγμα Ιεραρχιών	6
Σχήμα 2.2 Παράδειγμα star-schema	7
Σχήμα 2.3 Αποτελέσματα παραδείγματος για κάθε μήνα.....	13
Σχήμα 3.1 Η γενική μορφή ενός KPI σε SQL.....	20
Σχήμα 3.2 Πεδία για τη βάση δεδομένων.....	40
Σχήμα 3.3 Πεδίο Fact Table και φράγματα	40
Σχήμα 3.4 Πεδία για τη Μέτρηση	40
Σχήμα 3.5 Πίνακες διαστάσεων με τις παραμέτρους σαν φίλτρα	41
Σχήμα 3.6 UI Fact Table εμφάνιση Constraint Tables.....	42
Σχήμα 3.7 Database από Fact Table εμφάνιση Constraint Tables	43
Σχήμα 3.8 UI από Αποτελέσματα για τιμές αδέρφια.....	44
Σχήμα 3.9 Database Αποτελέσματα για τιμές αδέρφια.....	45
Σχήμα 3.10 UI από Αποτελέσματα ανάλυσης	46
Σχήμα 3.11 Database Αποτελέσματα με τιμή Store 4	47
Σχήμα 3.12 Database Αποτελέσματα με τιμή Store 18.....	47
Σχήμα 3.13 Database Αποτελέσματα με τιμή Store 12.....	47
Σχήμα 3.14 Αποτέλεσμα στο UI για σύγκριση πραγματικών με αναμενόμενες τιμές	48
Σχήμα 4.1 Package Diagram	51
Σχήμα 4.2 DatabaseController Package Diagram.....	51
Σχήμα 4.3 Class Diagram of package Models.....	52
Σχήμα 4.4 Class Diagram πακέτου Data Service	53
Σχήμα 4.5 Class Diagram όλου του πακέτου KPI Project	54
Σχήμα 5.1 Πίνακες πειραμάτων στη βάση Foodmart	57
Σχήμα 5.2 Σχεδιάγραμμα για χρόνο εκτέλεσης σε σχέση με τον αριθμό φίλτρων στη βάση foodmart.....	59
Σχήμα 5.3 Σχεδιάγραμμα για χρόνους εκτέλεσης σε σχέση με το πλήθος παρόμοιων τιμών (βάση foodmart).....	60
Σχήμα 5.4 Πίνακες πειραμάτων στη βάση pkdd99.....	61
Σχήμα 5.5 Σχεδιάγραμμα για χρόνο εκτέλεσης σε σχέση με τον αριθμό φίλτρων στη βάση pkdd99	62

Σχήμα 5.6 Σχεδιάγραμμα για χρόνους εκτέλεσης σε σχέση με το πλήθος παρόμοιων τιμών (βάση pkdd99).....	63
Σχήμα 5.7 Σχεδιάγραμμα για χρόνο εκτέλεσης σε σχέση με τον αριθμό φίλτρων στη βάση foodmart.....	65
Σχήμα 5.8 Σχεδιάγραμμα για χρόνο εκτέλεσης σε σχέση με τον αριθμό φίλτρων στη βάση για 100.000 εγγραφές.....	66
Σχήμα 5.9 Σχεδιάγραμμα για χρόνο εκτέλεσης σε σχέση με τον αριθμό φίλτρων στη βάση για 1.000.000 εγγραφές	66
Σχήμα 5.10 Σχεδιάγραμμα για χρόνο εκτέλεσης σε σχέση με τον αριθμό φίλτρων στη βάση pkdd99 για τις αναμενόμενες τιμές	68

ΚΑΤΑΛΟΓΟΣ ΑΛΓΟΡΙΘΜΩΝ

Αλγόριθμος 3.1 Παραγωγή αδερφών ερωτήσεων.....	23
Αλγόριθμος 3.2 Εύρεση τιμής ανωτέρου επιπέδου	24
Αλγόριθμος 3.3 Εύρεση παρόμοιων τιμών.....	24
Αλγόριθμος 3.4 Εύρεση τιμής ανωτέρου επιπέδου και συσχέτιση με άλλο πίνακα	25
Αλγόριθμος 3.5 Εύρεση παρόμοιων τιμών και συσχέτιση με άλλο πίνακα	25
Αλγόριθμος 3.6 Εκτέλεση όλων των σταδίων ενός KPI	26
Αλγόριθμος 3.7 Παραγωγή ερώτησης για παρόμοιες τιμές και μετρήσεις	27
Αλγόριθμος 3.8 Εύρεση τιμής ανωτέρου επιπέδου	28
Αλγόριθμος 3.9 Αποτελέσματα παρόμοιων τιμών με τις μετρήσεις.....	28
Αλγόριθμος 3.10 Εύρεση τιμής ανωτέρου επιπέδου με συσχέτιση σε άλλο πίνακα	29
Αλγόριθμος 3.11 Αποτελέσματα παρόμοιων τιμών με τις μετρήσεις	29
Αλγόριθμος 3.12 Εκτέλεση όλων των σταδίων ενός KPI	30
Αλγόριθμος 3.13 Σύγκριση της πραγματικής τιμής του KPI με την αναμενόμενη τιμή ως προς το ανώτερο επίπεδο.....	32
Αλγόριθμος 3.14 Εύρεση τιμής ανωτέρου επιπέδου.....	33
Αλγόριθμος 3.15 Υπολογισμός για το συνολικό ποσό για το έτος τους ανωτέρου επιπέδου	33
Αλγόριθμος 3.16 Υπολογισμός για το συνολικό ποσό για το έτος της παραμέτρου.....	34
Αλγόριθμος 3.17 Τύπος αναμενόμενης τιμής.....	34
Αλγόριθμος 3.18 Σύγκριση της πραγματικής τιμής του KPI με την αναμενόμενη τιμή ως προς το σύνολο	35
Αλγόριθμος 3.19 Υπολογισμός για το συνολικό ποσό για το έτος	36
Αλγόριθμος 3.20 Υπολογισμός για το συνολικό ποσό για το έτος	36
Αλγόριθμος 3.21 Τύπος αναμενόμενης τιμής για το σύνολο	36
Αλγόριθμος 3.22 Ερώτηση εύρεσης πινάκων συσχέτισης.....	37
Αλγόριθμος 3.23 Κώδικας παραγωγής της ερώτησης για τα αποτελέσματα-πίνακα συσχέτισης.....	38
Αλγόριθμος 3.24 Κώδικας παραγωγής της ερώτησης για τα αποτελέσματα-αδέρφια...	39

ΠΕΡΙΛΗΨΗ

Δήμητρα Γεωργιάδου, Δ.Μ.Σ. στη Μηχανική Δεδομένων και Υπολογιστικών Συστημάτων,
Τμήμα Μηχανικών Η/Υ και Πληροφορικής, Πολυτεχνική Σχολή, Πανεπιστήμιο
Ιωαννίνων, Ιανουάριος 2021

Αυτόματη παραγωγή αναφορών για την κατάσταση ενός Key Performance Indicator
Επιβλέπων : Παναγιώτης Βασιλειάδης - Καθηγητής

Οι Κρίσιμοι Δείκτες Απόδοσης (Key Performance Indicators -- KPIs) είναι σύνθετες μετρήσεις που χρησιμοποιούνται για την εκτίμηση της αποτελεσματικότητας ενός οργανισμού και την παρακολούθηση των επιδόσεων τους. Η παρούσα μεταπτυχιακή εργασία ειδίκευσης αναφέρεται στην υλοποίηση ενός ολοκληρωμένου συστήματος, το οποίο αφορά τη δημιουργία και αποτίμηση Κρίσιμων Δεικτών Απόδοσης. Ένας Κρίσιμος Δείκτης Απόδοσης είναι κατά βάση ένα συναθροιστικό ερώτημα σε μια βάση δεδομένων, το οποίο αποτιμά μια τιμή-μέτρηση σε σχέση με ένα υποσύνολο των δεδομένων, για το οποίο, ο αναλυτής έχει δώσει και ένα μέτρο της αναμενόμενης τιμής του, και ένα τρόπο χαρακτηρισμού του αποτελέσματος σε σχέση με την αναμενόμενη τιμή. Κατά τη δημιουργία ενός κρίσιμου δείκτη απόδοσης, ο χρήστης καλείται να δηλώσει τη βάση δεδομένων, για την οποία θα γίνει η ανάλυση, έναν κύριο πίνακα (Fact Table) και φίλτρα για τους πίνακες διαστάσεων με τους οποίους συσχετίζεται. Η αποτίμηση ενός KPI, αφορά τον υπολογισμό της τιμής του και την αυτόματη παραγωγή αναφορών με τα ευρήματα των αποτελεσμάτων, από τα δεδομένα που έχουν καταχωρηθεί. Για τη δημιουργία των αναφορών δημιουργούνται μια σειρά από ερωτήσεις, οι οποίες εκτελούνται στη βάση δεδομένων, ώστε να ανακτηθούν τα αποτελέσματα των μετρήσεων. Υπάρχουν δύο είδη αναφορών που συνοδεύουν ένα KPI. Το πρώτο είδος αναφοράς αποτιμά την τιμή του KPI σε σχέση με όμορα υποσύνολα δεδομένων – πρακτικά, για κάθε φίλτρο που έχει δηλώσει ο χρήστης, το σύστημα παράγει αυτόματα συγκρίσεις μεταξύ των παρόμοιων τιμών του. Το δεύτερο είδος αναφορών εκμεταλλεύεται την ύπαρξη χρονικών φίλτρων και εξάγει αναμενόμενες τιμές για ένα δείκτη που αφορά ένα συγκεκριμένο χρονικό σημείο, σε σχέση με τη συνολική τάση των δεδομένων σε σχέση με προηγούμενα χρονικά σημεία. Το σύστημα που κατασκευάστηκε απεικονίζει τα αποτελέσματα σε διαγράμματα και πίνακες. Οι προτεινόμενες μέθοδοι αποτιμώνται πειραματικά για την κλιμακωσιμότητά τους.

EXTENDED ABSTRACT

Dimitra Georgiadou, MSc in Computer Science, Department of Computer Science and Engineering, University of Ioannina, Greece, June 2022.

Automatic generation of reports at the status of a Key Performance Indicator.

Advisor: Panagiotis Vassiliadis, Professor.

Key Performance Indicators (KPIs) are complex metrics used to assess the effectiveness of an organization and monitor its performance. This Master's Thesis is about the implementation of an integrated system, which involves the creation and evaluation of Key Performance Indicators. A Key Performance Indicator is basically an aggregate query in a specific database, which evaluates a value-measure against a subset of data. The analyst has provided a measure of its expected value and a way of describing the results in terms of the expected value. At the creation of a key performance indicator, the user is asked to declare the database for which the analysis is to be performed, a master table (Fact Table), and filters for the dimension tables with which it is associated. The evaluation of a KPI involves calculating its value and automatically generating reports with the results from the data. To generate the reports, a series of queries are created and executed in the database to retrieve the results with the measures. There are two types of reports that are associated with a KPI. The first type of report evaluates the value of the KPI against similar subsets of data. For each filter declared by the user, the system automatically generates comparisons between similar values. The second type of reporting takes advantage of the existence of filters and extracts expected values for an indicator related to a specific point in time, relative to the overall trend of the data and relative to previous points in time. The system displays the results in graphs and tables. The proposed methods are evaluated experimentally for their scalability.

ΚΕΦΑΛΑΙΟ 1

ΕΙΣΑΓΩΓΗ

1.1 Στόχοι

1.2 Δομή της Διατριβής

1.1 Στόχοι

Η παρούσα μεταπτυχιακή εργασία ειδίκευσης έχει ως στόχο την υλοποίηση ενός ολοκληρωμένου συστήματος το οποίο θα διαθέτει μια βάση δεδομένων και τη διεπαφή του χρήστη, ώστε με τη δημιουργία Κρίσιμων Δεικτών Απόδοσης (KPIs) να μπορεί να κάνει μετρήσεις για την κατάσταση και την απόδοση του οργανισμού και αυτόματα να συνάγει επιπλέον μετρήσεις και αναφορές, που σχετίζονται με τον ορισθέντα δείκτη απόδοσης. Οι δείκτες απόδοσης είναι σύνθετες μετρήσεις που χρησιμοποιούνται για την εκτίμηση της αποτελεσματικότητας ενός οργανισμού και την παρακολούθηση των επιδόσεων των διαδικασιών και των επιχειρηματικών στρατηγικών του. Ορίζονται σε σχέση με μια επιχειρηματική στρατηγική και τους επιχειρησιακούς στόχους, παρέχοντας μια συνολική επισκόπηση της κατάστασης της επιχείρησης. Συνήθως περιλαμβάνονται σε πίνακες και αναφορές, παρέχοντας μια λεπτομερή απεικόνιση για κάθε τομέα του οργανισμού. Έτσι, οι χρήστες μπορούν να αξιολογούν και να διαχειριστούν την απόδοση με τη χρήση των δεικτών, για την υποστήριξη της λήψης αποφάσεων. Οι δείκτες έχουν συνήθως μια τρέχουσα τιμή, η οποία προκύπτει από την εκτέλεση μίας (ή περισσότερων) ερώτησης σε μια βάση δεδομένων με τις σχετικές πληροφορίες και συγκρίνεται με μια τιμή του στόχου, μια τιμή κατωφλίου και μια ελάχιστη τιμή.

Στο πλαίσιο της παρούσας μεταπτυχιακής εργασίας υλοποιήθηκε σύστημα με το οποίο οι αναλυτές μπορούν να καταχωρούν KPIs' που τους ενδιαφέρουν και να λαμβάνουν μια σφαιρική ενημέρωση για την κατάσταση τους. Συγκεκριμένα το σύστημα που υλοποιήθηκε έχει τις εξής λειτουργίες:

- Εγγραφή ενός νέου KPI πάνω σε μια OLAP βάσης κύβων, με τα μέτρα και τις ιεραρχίες για τη διαχείριση ή επερώτηση της πληροφορίας. Κατά τη δημιουργία ενός KPI ο χρήστης καταχωρεί βάση δεδομένων, κύριο πίνακα ανάλυσης (Fact Table), φίλτρα στους πίνακες διαστάσεων με τους οποίους συσχετίζεται, μέτρηση και φράγματα.
- Για την παρακολούθηση της απόδοσης του οργανισμού ο χρήστης μπορεί να την ελέγξει από δύο είδη αναφορών που προκύπτουν από μια σειρά ερωτήσεων προς τη βάση δεδομένων. Τα είδη των αναφορών είναι τα εξής:
 - Στην πρώτη αναφορά για κάθε φίλτρο που έχει δηλώσει ο χρήστης στους πίνακες διαστάσεων θα βρεθούν και οι παρόμοιες τιμές και θα υπολογιστούν τα αποτελέσματα του συναθροιστικού ποσού της μέτρησης, ώστε να γίνουν οι μεταξύ τους συγκρίσεις. Τα αποτελέσματα απεικονίζονται σε διαγράμματα καθώς και σε πίνακες για κάθε παρόμοια τιμή του φίλτρου με το ποσό της συναθροιστικής τιμής της μέτρησης.
 - Στη δεύτερη αναφορά γίνεται σύγκριση ως προς την αναμενόμενη τιμή ενός φίλτρου και την πραγματική τιμή. Η αναμενόμενη τιμή υπολογίζεται για κάθε φίλτρο των διαστάσεων ως προς το ανώτερο επίπεδο και ως προς το σύνολο εγγραφών. Τα αποτελέσματα απεικονίζονται σε διαγράμματα καθώς και σε πίνακες, ώστε να γίνουν οι συγκρίσεις των αναμενόμενων τιμών και των πραγματικών.

1.1 Δομή Διατριβής

Τα κεφάλαια από τα οποία απαρτίζεται η εργασία είναι τα εξής:

- Κεφάλαιο 2: αναγράφεται το θεωρητικό υπόβαθρο που απαιτείται για την υλοποίηση της εργασίας καθώς και ορισμοί που χρησιμοποιούνται στα υπόλοιπα κεφάλαια.
- Κεφάλαιο 3: καταγράφονται όλοι οι αλγόριθμοι που υλοποιήθηκαν για την ανάλυση και τις συγκρίσεις ενός KPI

- Κεφάλαιο 4: σημειώνονται όλα τα προγραμματιστικά εργαλεία που χρησιμοποιήθηκαν καθώς και η αρχιτεκτονική του λογισμικού
- Κεφάλαιο 5: καταγράφονται όλα τα πειράματα που διεξήχθησαν για τις υλοποιήσεις

ΚΕΦΑΛΑΙΟ 2

ΘΕΩΡΗΤΙΚΟ ΥΠΟΒΑΘΡΟ

-
- 2.1 Data warehouses
 - 2.2 Κύβος
 - 2.3 Διαστάσεις
 - 2.4 Ιεραρχίες
 - 2.5 Star-schema
 - 2.6 Βασικός δείκτης απόδοσης
 - 2.7 KPIs in Analysis Services
-

Στην παρούσα ενότητα παραθέτουμε τη σχετική ορολογία και τεχνολογίες. Ο αναγνώστης παραπέμπεται στο [VAZI14] για μεγαλύτερη εμβάθυνση.

2.1 Data warehouses

Οι DWs υπάρχουν για το συνδυασμό δύο σκοπών: (α) να παρέχουν μια ολοκληρωμένη προβολή όλων των δεδομένων των διαφόρων υποσυστημάτων, με εγγυήσεις ποιότητας και ακεραιότητας των δεδομένων, (β) να αποφορτίζουν την επεξεργασία σύνθετων ερωτημάτων από τα συστήματα παραγωγής.

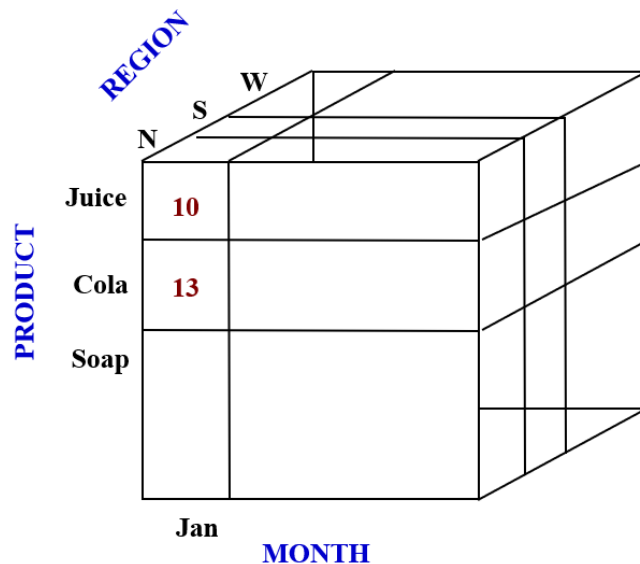
Υπάρχει μεγάλη ποικιλία στην τεχνολογία και στα εργαλεία για την πρόσβαση, ανάλυση, εξαγωγή και αξιοποίηση των δεδομένων που περιέχονται στις αποθήκες δεδομένων (data warehouses). Ο τυπικός μηχανισμός OLAP (on line analytical processing) επιτρέπει στους χρήστες να αλληλεπιδρούν ώστε να υποβάλουν ερωτήματα σε μια αποθήκη δεδομένων. Τα Data mining Εργαλεία χρησιμοποιούνται για την εξαγωγή συμπερασμάτων. Η αποθήκη δεδομένων μπορεί να συνοψιστεί ως προς:

- Αναφορές (Reporting) σαν πίνακες ελέγχου

- Διαχείριση επίδοσης, με μετρικές, βασικούς δείκτες αποδόσεων (KPIs) και κάρτες αποτελεσμάτων
- Ανάλυση, όπως είναι η OLAP, εξόρυξη δεδομένων, ανάλυση χρονοσειρών, εξόρυξη κειμένου, web analytics και προηγμένη οπτικοποίηση δεδομένων.

2.2 Κύβος (Cube)

Οι αποθήκες δεδομένων και τα συστήματα OLAP βασίζονται στο πολυδιάστατο μοντέλο, το οποίο έχει δεδομένα σε n-δισδιάστατο χώρο που ονομάζεται κύβος. Χρησιμοποιείται για την ανάλυση δεδομένων. Ορίζεται από διαστάσεις (dimensions) και γεγονότα (facts) ή κελιά (cells) τα οποία σχετίζονται από τιμές που ονομάζονται μέτρα (measures). Τα μέτρα αυτά χρησιμοποιούνται για την αξιολόγηση της ανάλυσης.



Σχήμα 2.1 Ένας τρισδιάστατος κύβος για δεδομένα πωλήσεων με διαστάσεις Product, Region και Month¹

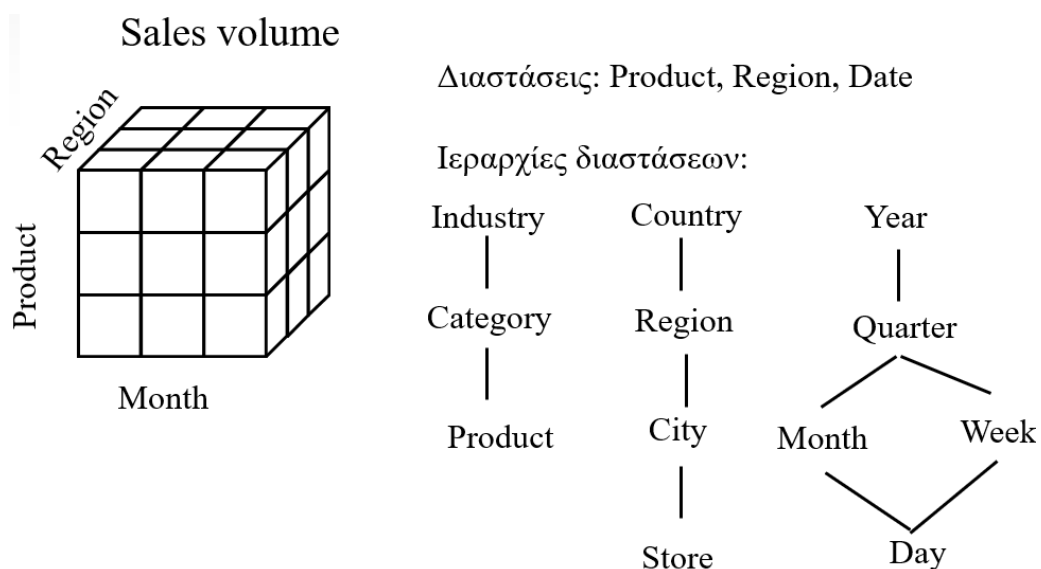
¹ Με έγκριση του συγγραφέα

2.3 Διαστάσεις (Dimensions)

Οι διαστάσεις είναι ορισμοί που χρησιμοποιούνται για την ανάλυση των δεδομένων του κύβου. Ένα επίπεδο διάστασης αναπαριστά το επίπεδο της λεπτομέρειας για την κάθε διάσταση του κύβου.

2.4 Ιεραρχίες (Hierarchies)

Προκειμένου να γίνει αντιληπτή και να εξαχθεί η γνώση από έναν κύβο πρέπει να δούμε τα δεδομένα με λεπτομέρεια σε διάφορα επίπεδα. Οι Ιεραρχίες επιτρέπουν αυτή τη δυνατότητα με την έννοια μιας ακολουθίας αντιστοιχίσεων που αφορούν τις λεπτομέρειες χαμηλότερου επιπέδου με αυτές του υψηλότερου επιπέδου, σε πιο γενικά σενάρια.

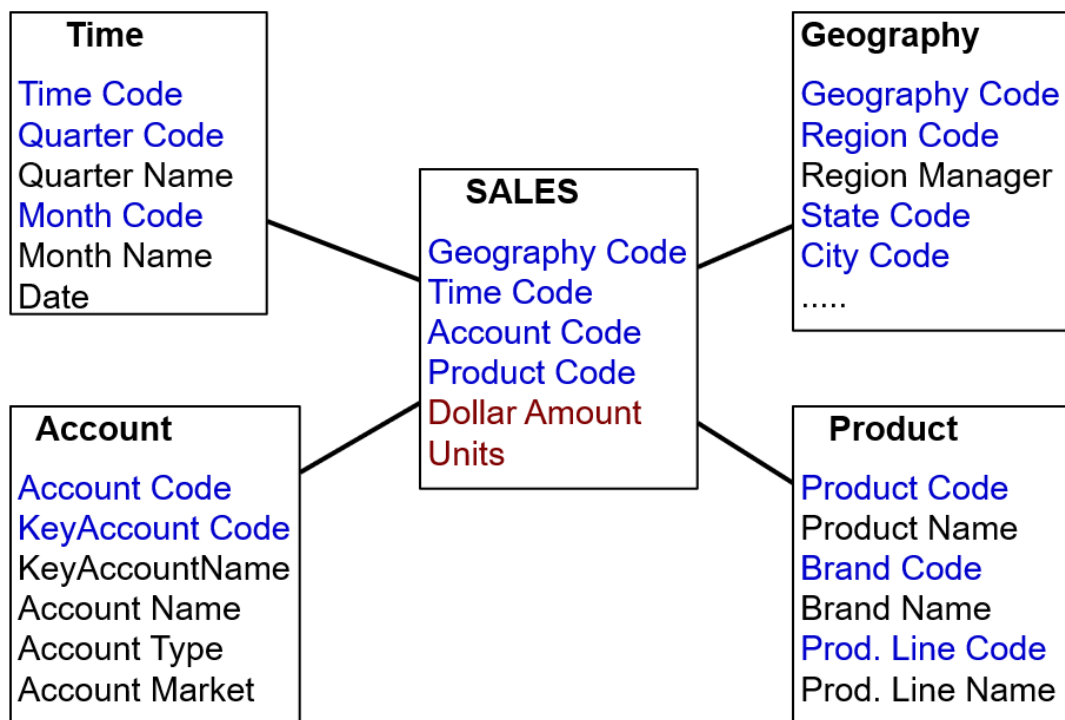


Σχήμα 2.1 Παράδειγμα Ιεραρχιών²

² Με έγκριση του συγγραφέα

2.5 Star-schema

Με τον όρο star-schema εννοούμε μια αναπαράσταση του πολυδιάστατου μοντέλου σε ένα σχεσιακό DBMS, όπου υπάρχει ένας κεντρικός πίνακας και ένα σύνολο από πίνακες διαστάσεων για κάθε μια διάσταση. Ο κεντρικός πίνακας περιέχει τα ξένα κλειδιά του συνόλου από τους πίνακες διαστάσεων. Οι πίνακες δεν είναι κανονικοποιημένοι και περιέχουν πολλά περιττά δεδομένα, πιο έντονο είναι αυτό με την παρουσία των ιεραρχιών.



Σχήμα 2.2 Παράδειγμα star-schema³

Όλα τα παραπάνω σε συνδυασμό μεταξύ τους είναι η βάση για την υλοποίηση του δείκτη απόδοσης. Με βάση τα παραπάνω θα γίνεται η δημιουργία, επεξεργασία ανάκτηση όλη της πληροφορίας καθώς και η συσχέτιση των δεδομένων που αφορούν έναν οργανισμό σε συνδυασμό με το δείκτη απόδοσης, ώστε να εξάγονται οι πληροφορίες της απόδοσης του οργανισμού. Τα αποτελέσματα θα αναπαρίστανται σε μια αναφορά για να γίνεται η πληροφορία διαδραστική και κατανοητή από τον χρήστη.

³ Με έγκριση του συγγραφέα από Stanford Technology Group, Inc., 1996

2.6 Βασικός δείκτης απόδοσης *Mg{ Rgt hto cpeg Kpf lec vqt +

Συχνά οι οργανισμοί προκειμένου να εκτιμήσουν κάποιες στατιστικές μετρήσεις, ώστε να δείξουν την απόδοση τους ή συγκρίσεις μεταξύ των προϊόντων τους ή ακόμα σύγκριση μεταξύ χρονικών περιόδων χρησιμοποιούν κάποιους δείκτες. Οι δείκτες αυτοί ονομάζονται βασικοί δείκτες απόδοσης (KPIs). Ορίζονται σε σχέση με τους στόχους που έχει θέση μια επιχείρηση ή ένας οργανισμός ώστε να γίνεται επισκόπηση της κατάστασης. Η αναπαράσταση γίνεται συνήθως με αναφορές (reports) με λεπτομερή απεικόνιση των δεδομένων για κάθε τομέα. Για να γίνει αυτή η σύγκριση οι χρήστες έχουν μια τρέχουσα τιμή και γίνεται σύγκριση με τον στόχο μια τιμή κατωφλίου και μια ελάχιστη τιμή. Οι δείκτες KPI μπορούν να ταξινομηθούν κατά μήκος άλλων διαστάσεων για παράδειγμα η χρονική διάσταση:

- Leading KPIs: προσδοκίες ως προς το τι μπορεί να συμβεί στο μέλλον
- Coincident KPIs: αντικατοπτρίζουν αυτό που συμβαίνει σήμερα
- Lagging KPIs: αντικατοπτρίζουν τι συνέβη στο παρελθόν
- Input KPIs: μετρούν τους πόρους που επενδύονται ή χρησιμοποιούνται στην παραγωγή
- Output KPI: αντικατοπτρίζουν τα συνολικά αποτελέσματα
- Qualitative KPI: οι οποίοι μετρούν ένα περιγραφικό χαρακτηριστικό, μια γνώμη, ή μια ιδιότητα
- Quantitative KPIs: οι οποίοι μετρούν χαρακτηριστικά που λαμβάνονται μέσω μιας μαθηματικής έκφρασης
- Strategic KPIs: οι οποίοι συνήθως αναφέρονται σε ανώτερα επίπεδα του οργανισμού και σε λιγότερο τακτά χρονικά διαστήματα από τους αντίστοιχους επιχειρησιακούς δείκτες. Έχουν μεσοπρόθεσμη ή μακροπρόθεσμη χρονική εμβέλεια.
- Operational KPIs: οι οποίοι επικεντρώνονται σε χαμηλότερα επίπεδα του οργανισμού
- και αναφέρονται συχνότερα από τους στρατηγικούς δείκτες. Συνήθως
- έχουν βραχυπρόθεσμη χρονική εμβέλεια.

Οδηγός ορισμού δείκτη KPI. Για τον καλύτερο ορισμό των δεικτών KPI για έναν οργανισμό θα πρέπει να προσδιοριστούν οι πηγές από τις οποίες μπορεί να αντληθεί η πληροφορία. Αυτές οι πηγές ταξινομούνται σε πρωτογενείς, δευτερογενείς και εξωτερικές.

- Πρωτογενείς πηγές: εδώ βρίσκονται οι εργαζόμενοι πρώτης γραμμής με την έννοια ότι είναι αυτοί που γνωρίζουν ποιοι είναι οι βασικοί παράγοντες για την επίτευξη των στόχων του οργανισμού
 - Διευθυντές (Managers): παρέχουν τις γνώσεις για την στρατηγική
 - Διοικητικό Συμβούλιο: καθορίζει τους οργανωτικούς στόχους και προτείνει συγκεκριμένους δείκτες μέτρησης της απόδοσης (KPIs) που έχουν υψηλή προτεραιότητα και κάποιες φορές δεν διαπραγματεύονται
 - Προμηθευτές και πελάτες: έχουν την εξωτερική προοπτική ως προς αυτό που πρέπει να βελτιωθεί
 - Δευτερογενείς πηγές: περιλαμβάνουν το στρατηγικό σχέδιο ανάπτυξης, το ετήσιο επιχειρηματικό/στρατηγικό σχέδιο, ετήσιες εκθέσεις, εσωτερικές επιχειρησιακές εκθέσεις και εκθέσεις επισκόπησης των ανταγωνιστών
 - Εξωτερικές πηγές: περιλαμβάνουν συμβουλές εμπειρογνομόνων
- Μόλις οριστούν οι πηγές ακολουθούν τα παρακάτω βήματα:
1. Συγκέντρωση (μικρής) ομάδας
 2. Κατηγοριοποίηση πιθανών δεικτών
 3. Ιδέες και συζήτηση για πιθανά μέτρα πριν το τελικό σύνολο αποφασιστεί
 4. Προτεραιότητα στις μετρήσεις που έχουν καθοριστεί αρχικά
 5. Εκτέλεση των τελικών φίλτρων. Αυτό σημαίνει ότι ορισμός της μετρικής είναι σαφής και ξεκάθαρος για τα άτομα που δεν ανήκουν στην βασική ομάδα, ότι έχουμε αξιόπιστα δεδομένα για τον υπολογισμό της μετρικής και να βεβαιωθούμε ότι η επίτευξη των μετρικών θα οδηγήσει στην επίτευξη των στόχων μας
 6. Ορισμός στόχων για τις επιλεγμένες μετρικές

2.7 KPIs in Analysis Services

Οι αλγόριθμοι εξόρυξης δεδομένων δημιουργούν ένα μοντέλο εξόρυξης αναλύοντας ένα σύνολο δεδομένων για την εύρεση κρυφών μοτίβων και τάσεων. Μόλις καθοριστούν οι παράμετροι του μοντέλου εξόρυξης, εφαρμόζονται σε ολόκληρο το σύνολο των δεδομένων για την εξαγωγή ενδιαφέρουσας γνώσης. Το μοντέλο εξόρυξης μπορεί να λάβει διάφορες μορφές, όπως ένα δέντρο αποφάσεων που προβλέπει αν ένας συγκεκριμένος πελάτης θα αγοράσει ένα προϊόν, ένα μοντέλο που προβλέπει πωλήσεις, ένα σύνολο συστάδων που περιγράφουν τα προφίλ των πελατών κ.ο.κ. Τα δύο εργαλεία που χρησιμοποιούμε για να υλοποιήσουμε τα παραδείγματά μας, είναι: Microsoft

Analysis Services και Pentaho Business Analytics, παρέχουν δεδομένα δυνατότητες εξόρυξης δεδομένων που υλοποιούν τις έννοιες που μελετήθηκαν παραπάνω. Το πρώτο παρέχει επίσης μια γλώσσα ερωτημάτων με την ονομασία DMX (Data Mining eXtensions). Η γλώσσα αυτή μπορεί να χρησιμοποιηθεί για την υποβολή ερωτημάτων σε μοντέλα εξόρυξης δεδομένων στο Analysis Services. Υπάρχουν δύο είδη ερωτημάτων DMX: ερωτήματα περιεχομένου, τα οποία παρέχουν λεπτομέρειες σχετικά με τα πρότυπα που ανακαλύφθηκαν κατά την ανάλυση, και ερωτήματα πρόβλεψης, τα οποία χρησιμοποιούν τα μοτίβα του μοντέλου, για παράδειγμα, για να κάνουν προβλέψεις για νέα δεδομένα. Από την άλλη πλευρά, το Pentaho Data Mining είναι βασίζεται στο Weka (Waikato Environment for Knowledge Analysis). Το Weka είναι ένα λογισμικό μηχανικής μάθησης και εξόρυξης δεδομένων γραμμένο σε Java. Στη συνέχεια μελετάμε τρεις αλγόριθμους που παρέχονται από τις Analysis Service, συγκεκριμένα, τα δέντρα αποφάσεων, την ομαδοποίηση και τους κανόνες συσχέτισης.

– Δέντρα απόφασης (Decision Trees)

Ένα μοντέλο δέντρου αποφάσεων είναι ένας αλγόριθμος που απαιτεί τα χαρακτηριστικά της κλάσης που πρόκειται να προβλεφθεί, τα χαρακτηριστικά που πρέπει να χρησιμοποιηθούν ως είσοδος και τα χαρακτηριστικά που θα αγνοηθούν, από τον αλγόριθμο αλλά μπορούν να χρησιμοποιηθούν για την οπτικοποίηση των αποτελεσμάτων. Δεδομένου ενός μοντέλου δέντρου αποφάσεων, ένα ερώτημα περιεχομένου DMX πάνω σε αυτό μπορεί να παρέχει στατιστικά στοιχεία σχετικά με τον αριθμό των περιπτώσεων σε κάθε επίπεδο του δέντρου ή τους κανόνες σε κάθε κόμβο, ενώ ένα ερώτημα πρόβλεψης αντιστοιχίζει το μοντέλο σε νέα δεδομένα προκειμένου να παράγει συστάσεις και ταξινομήσεις. Τα δεδομένα σχετικά με το μοντέλο μπορούν επίσης να ανακτηθούν. Για παράδειγμα, πάνω από το μοντέλο που αναπτύσσεται (το οποίο ονομάζεται NWDW Decision Tree), μπορούμε να εκδώσουμε το ακόλουθο ερώτημα περιεχομένου ζητώντας τα χαρακτηριστικά ενός εσωτερικού κόμβου στο δέντρο.

– Ομαδοποίηση (Clustering)

Στα μοντέλα ομαδοποίησης, ένα ερώτημα περιεχομένου (query) ζητά λεπτομέρειες σχετικά με τις ομάδες που βρέθηκαν. Ένα ερώτημα πρόβλεψης μπορεί να ρωτά σε ποια συστάδα ένα νέο δεδομένο σημείο είναι πιθανότερο να ανήκει. Αφού κατασκευαστεί το μοντέλο (NWDW Clustering), μπορούν να βρεθούν τα χαρακτηριστικά των παραγόμενων συστάδων. Δεδομένου ότι στις υπηρεσίες Analysis Services, η δομή συσταδοποίησης είναι ένα δέντρο, έτσι ώστε κάτω από τη ρίζα να υπάρχει μια συλλογή επίπεδων κόμβων.

Συνεπώς, εφόσον όλες οι συστάδες έχουν τύπο κόμβου 5, μπορεί εύκολα να ανακτηθεί ένας κατάλογος των συστάδων με την αναζήτηση του περιεχομένου του μοντέλου μόνο για τους κόμβους αυτού του τύπου. Το μοντέλο αυτό μπορεί να χρησιμοποιηθεί για προβλέψεις σχετικά με το αποτέλεσμα χρησιμοποιώντας τα προβλέψιμα χαρακτηριστικά στο μοντέλο.

– Κανόνες συσχέτισης (Association Rules)

Υπάρχουν δύο κοινές χρήσεις ενός μοντέλου συσχέτισης: για την ανακάλυψη πληροφοριών σχετικά με τα συχνά σύνολα στοιχείων και για την εξαγωγή λεπτομερειών σχετικά με συγκεκριμένους κανόνες και σύνολα στοιχείων. Για παράδειγμα, μπορεί να ανακτηθεί μια λίστα κανόνων που βαθμολογήθηκαν ως ενδιαφέροντες ή να δημιουργηθεί μια λίστα με τα πιο συχνά σύνολα αντικειμένων. Αυτές οι πληροφορίες μπορούν να ληφθούν μέσω ενός ερωτήματος περιεχομένου DMX ή μέσω περιήγησης αυτές τις πληροφορίες με τη χρήση του προγράμματος προβολής συσχετίσεων της Microsoft..

Τα Analysis Services παρέχουν ένα framework για τον ορισμό KPIs που εκμεταλλεύονται τις Δεδομένα της επιχείρησης που είναι αποθηκευμένα σε κύβους. Κάθε KPI διαθέτει ένα προκαθορισμένο σύνολο χαρακτηριστικών και ανατίθενται εκφράσεις multidimensional expressions (MDX). Μόνο τα metadata για τα KPIs αποθηκεύονται από τις υπηρεσίες Analysis Services, οι συναρτήσεις των MDX είναι διαθέσιμες σε εφαρμογές ώστε να μπορεί να γίνει η ανάκτηση τιμών KPI από κύβους χρησιμοποιώντας τα metadata.

Στο Analysis Services, ένας κύβος μπορεί να έχει μια συλλογή από KPIs. Κάθε KPI έχει πέντε ιδιότητες, οι οποίες είναι εκφράσεις MDX που επιστρέφουν αριθμητικές τιμές από έναν κύβο, όπως περιγράφεται στη συνέχεια:

- Value, η οποία επιστρέφει την πραγματική τιμή του KPI. Είναι υποχρεωτική για έναν KPI.
- Goal, η οποία επιστρέφει τον στόχο του KPI.
- Status, που επιστρέφει την κατάσταση του KPI (μια τιμή μεταξύ -1 και 1)
- Trend, η οποία επιστρέφει το trend του KPI με την πάροδο του χρόνου. Επιστρέφει μια τιμή μεταξύ -1 και 1.

-Weight, επιστρέφει το βάρος του KPI. Εάν ένας KPI έχει ένα γονικό KPI, μπορούμε να ορίσουμε βάρη για να ελέγξουμε τη συμβολή αυτού του KPI στον γονέα του.

Οι υπηρεσίες ανάλυσης δημιουργούν κρυφά μέλη στις μετρήσεις (measures) διαστάσεις (dimensions) για κάθε ιδιότητα του KPI. Παρ' όλα αυτά, αυτά τα υπολογιζόμενα μέτρα μπορούν να χρησιμοποιηθούν σε μια έκφραση MDX, παρόλο που είναι κρυμμένα.

Ένα παράδειγμα για το πώς ο KPI δείκτης απόδοσης μπορεί να υλοποιηθεί με τη χρήση Analysis Services είναι το παρακάτω. Θέλουμε να παρακολουθούμε το ύψος των πωλήσεων ανά έτος σε σχέση με τον στόχο της επίτευξης του 15% αύξησης από έτος σε έτος. Αν το πραγματικό ποσό των πωλήσεων είναι μεγαλύτερο από το 95% του στόχου, η απόδοση είναι θεωρείται ικανοποιητική, αν, ωστόσο, το ποσό των πωλήσεων είναι εντός του 85-95% του στόχου, η διοίκηση πρέπει να ειδοποιηθεί, αν πέσει κάτω από το 85% του του στόχου, η διοίκηση πρέπει να λάβει άμεσα μέτρα για την αλλαγή της τάσης. Αυτές οι ειδοποιήσεις για δράση συνδέονται συνήθως με τη χρήση των δεικτών απόδοσης. Το ενδιαφέρον εστιάζεται στις τάσεις που σχετίζονται με το ποσό των πωλήσεων, αν το ποσό των πωλήσεων είναι 20% υψηλότερο από το αναμενόμενο, θα πρέπει να επισημανθεί. Ομοίως, αν το ποσό των πωλήσεων είναι 20% χαμηλότερο από το αναμενόμενο, τότε πρέπει να αντιμετωπίσουμε άμεσα την κατάσταση. Το ερώτημα MDX που υπολογίζει τον στόχο του KPI δίνεται στη συνέχεια:

WITH MEMBER Measures.SalesPerformanceGoal AS

CASE

WHEN ISEMPTY(PARALLELPERIOD(
[Order Date].Calendar.Month, 12,
[Order Date].Calendar.CurrentMember))

THEN Measures.[Sales Amount]

ELSE 1.15 *

(Measures.[Sales Amount],
PARALLELPERIOD (
[Order Date].Calendar.Month, 12,
[Order Date].Calendar.CurrentMember))

END,

FORMAT STRING = '\$###,##0.00'

```
SELECT { [Sales Amount], SalesPerformanceGoal } ON COLUMNS,
      [Order Date].Calendar.Month.MEMBERS ON ROWS
FROM Sales
```

Στο παραπάνω ερώτημα, η δήλωση CASE θέτει ως στόχο τις πωλήσεις του τρέχοντος μήνα, εάν δεν περιλαμβάνεται ο αντίστοιχος μήνας του προηγούμενου έτους στο χρονικό πλαίσιο του κύβου. Αυτό το ερώτημα δίνει το ακόλουθο αποτέλεσμα:

	Sales Amount	SalesPerformanceGoal
...	...	...
June 1997	\$33,843.80	\$33,843.80
July 1997	\$51,020.86	\$32,041.18
August 1997	\$45,841.67	\$29,308.07
September 1997	\$50,105.74	\$25,270.56
October 1997	\$62,651.25	\$40,801.50

Σχήμα 2.3 Αποτελέσματα παραδείγματος για κάθε μήνα

Όπως φαίνεται παραπάνω, από τότε που οι πωλήσεις στην αποθήκη δεδομένων της Northwind άρχισαν να τον Ιούλιο του 1996, ο στόχος έως τον Ιούνιο του 1997 έχει οριστεί στις τρέχουσες πωλήσεις. Μπορούμε να χρησιμοποιήσουμε τα εργαλεία δεδομένων του SQL Server Data Tools για τον ορισμό του παραπάνω KPI, τον οποίο ονομάζουμε Sales Performance (Απόδοση πωλήσεων). Για το σκοπό αυτό, πρέπει να παρέχουμε εκφράσεις MDX για κάθε μία από τις παραπάνω ιδιότητες ως εξής:

- Τιμή: Το μέτρο που χρησιμοποιείται για τον ορισμό του KPI είναι το [Measures].[SalesAmount].
- Στόχος: Ο στόχος για αύξηση 15% σε σχέση με το ποσό των πωλήσεων του προηγούμενου έτους δίνεται από την CASE έκφραση στο παραπάνω ερώτημα.
- Κατάσταση: Πρέπει να επιλέξουμε έναν γραφικό δείκτη για την εμφάνιση της κατάστασης του KPI.

2.8 Μελέτη UX KPI με βάση το ερωτηματολόγιο εμπειρίας χρήστη

Στη μελέτη [HIN+19] οι συγγραφείς παρουσιάζουν μια μέθοδο για τον υπολογισμό του user experience key performance indicator (UX KPI) χρησιμοποιώντας ένα νέο ερωτηματολόγιο της εμπειρίας του χρήστη (UX). Σε πολλές εταιρείες, οι αποφάσεις λαμβάνονται με βάση τα αριθμητικά στοιχεία (key figures), όπως ο κύκλος εργασιών, το κέρδος ή την ικανοποίηση των εργαζομένων. Οι μάνατζερ έχουν συνηθίσει οι πληροφορίες να συνοψίζονται και να είναι διαθέσιμες μέσω βασικών αριθμητικών στοιχείων. Κάθε αριθμητικό στοιχείο αντιπροσωπεύει μια πτυχή που είναι σημαντική για την επιτυχία της επιχείρησης ή των προϊόντων της. Για αυτό λόγο, κάθε τμήμα που συμβάλλει στην επιτυχία ενός προϊόντος έχει συνήθως με μεμονωμένα key figures που αντικατοπτρίζουν τη συμβολή του τμήματος. Η παρακολούθηση μπορεί να πραγματοποιηθεί με βάση αυτά τα βασικά μεγέθη για να γίνει καλύτερη η καθοδήγηση της εταιρείας. Αυτό οδηγεί στην απαίτηση να αποτυπωθούν τα αποτελέσματα των μετρήσεων που αφορούν το ερωτηματολόγιο (UX) σε έναν ενιαίο αριθμό, δηλαδή σε έναν βασικό δείκτη απόδοσης (KPI). Στο Ερωτηματολόγιο Εμπειρίας Χρήστη (UEQ), οι συγγραφείς ερεύνησαν τις απόψεις των ερωτηθέντων σχετικά με τη σημασία να υπάρχουν και κλίμακες στο ερωτηματολόγιο (UEQ), και για αυτό πρόσθεσαν άλλες έξι ερωτήσεις. Το UX KPI που προκύπτει συνοψίζει την υποκειμενική βαθμολογία ενός προϊόντος ώστε να υπάρχουν κλίμακες μέσα στο UEQ KPI. Οι συγγραφείς μπόρεσαν να αξιολογήσουν αυτή τη μέθοδο σε μια αρχική μελέτη σε τρεις χώρες (Αγγλία, Γερμανία και Ισπανία) με συνολικά 882 συμμετέχοντες αξιολογώντας το Amazon και το Skype. Το UEQ έχει επεκταθεί με έξι επιπλέον ερωτήσεις, αλλά ένα πλήρως συμπληρωμένο ερωτηματολόγιο εξακολουθεί να είναι διαθέσιμο. Αυτό σημαίνει ότι οι πληροφορίες για τις επιμέρους κλίμακες εξακολουθούν να είναι διαθέσιμες. Όποιος είναι υπεύθυνος για ένα προϊόν έχει την επιλογή, από τη μία πλευρά, μπορεί να χρησιμοποιήσει όλες τις πληροφορίες που προσφέρει το UEQ, από την άλλη πλευρά, μπορεί να χρησιμοποιήσει το UEQ KPI για να αναφέρει ένα έγκυρο αριθμητικό στοιχείο στην εταιρεία.

2.9 Μελέτη ανάλυσης επιχειρηματικών διαδικασιών με αυτόματο εντοπισμό κατωφλίων KPI

Το άρθρο [Abe16] περιγράφει μια μέθοδο για την αυτόματη ανίχνευση κατωφλίων των βασικών δεικτών απόδοσης (KPI) με τη διαίρεση και τη συνάθροιση περιπτώσεων διεργασιών σύμφωνα με τις διαφορές στα μοντέλα διεργασιών. Τα κατώφλια καθορίζουν τις αποδεκτές υψηλές και χαμηλές τιμές για τα δεδομένα που συλλέγονται. Τα κατώφλια μπορούν να χρησιμοποιηθούν ως ανάλυση του άξονα διερεύνησης δεδομένων για τα μοντέλων διεργασιών που ανακαλύπτονται από τα τεράστια αρχεία καταγραφής. Η προτεινόμενη μέθοδος επιτρέπει στους χρήστες να ελαχιστοποιήσουν το χρόνο που απαιτείται για την ανίχνευση κατωφλίων KPI μέσω δοκιμής και σφάλματος. Οι συγγραφείς εφάρμοσαν τη μέθοδο σε πραγματικά αρχεία καταγραφής και τα αποτελέσματα του πειράματος, έδειξαν ότι τα κατώφλια εντοπίστηκαν για δύο τύπους KPI. Παρόλο που ο ένας τύπος δεν συσχετιζόταν με μοτίβα διεργασιών, ο άλλος συσχετιζόταν σε μεγάλο βαθμό με αυτά. Τέτοια ευρήματα προκύπτουν συνήθως από τη γνώση του τομέα των επιχειρησιακών χρηστών και τα αποτελέσματα της ανάλυσης που διαθέτουν οι αναλυτές δεδομένων με τεχνικές γνώσεις. Η προσέγγισή αυτή για τα κατώφλια μπορεί να τα εντοπίζει αυτόματα και αυτό βοηθά στην επέκταση της ανάλυσης διαδικασιών για τους τελικούς χρήστες.

2.10 Ενίσχυση της λήψης αποφάσεων μέσω δια δραστικής ανάλυσης What-If

Στη μελέτη [Gat+22], δίνεται ο ορισμός του θεμελιώδη στόχου της ανάλυσης επιχειρηματικών δεδομένων, είναι η βελτίωση των επιχειρηματικών αποφάσεων με τη χρήση δεδομένων. Αυτός ο στόχος, απαιτεί κατανόηση και διερεύνηση των σχέσεων μεταξύ μεταβλητών δεδομένων που αντιπροσωπεύουν τους οδηγούς υπόθεσης και εκείνων που αντιπροσωπεύουν τους βασικούς δείκτες απόδοσης (KPIs), για τους οποίους ενδιαφέρονται οι χρήστες των επιχειρήσεων. Οι χρήστες συχνά λαμβάνουν αποφάσεις για να επιτύχουν βασικούς δείκτες απόδοσης (KPIs), όπως η αύξηση της διατήρησης πελατών ή των πωλήσεων ή η μείωση του κόστους. Για την ανακάλυψη της σχέσης μεταξύ των χαρακτηριστικών δεδομένων που αποτελούν οδηγούς και εκείνων που αντιστοιχούν στους KPIs ενδιαφέροντος, οι χρήστες των επιχειρήσεων πρέπει να εκτελούν μακροχρόνιες διερευνητικές αναλύσεις. Οι συγγραφείς παρουσιάζουν τέσσερις

λειτουργίες για τη διαδραστική εκμάθηση και το συλλογισμό σχετικά με αυτές τις σχέσεις. Η επιτυχία των αναλύσεων με βάση το μοντέλο για το ευρύτερο κοινό χωρίς υπόβαθρο στην κωδικοποίηση ή την επιστήμη των δεδομένων εξαρτάται από την κατανόηση που διαθέτουν οι χρήστες, τους στόχους τους και τον ανάλογο σχεδιασμό της εμπειρίας.

Οι συγγραφείς υλοποίησαν αυτές τις λειτουργίες στο SYSTEMD, σε ένα διαδραστικό σύστημα οπτικής ανάλυσης δεδομένων, που επιτρέπει στους χρήστες των επιχειρήσεων να πειραματιστούν με τα δεδομένα θέτοντας ερωτήσεις whatif. Οι συγγραφείς αξιολόγησαν το σύστημα μέσω τριών περιπτώσεων: μοντελοποίηση συνδυασμού μάρκετινγκ, ανάλυση διατήρησης πελατών και ανάλυση κλεισίματος συμφωνιών και αναφέρουν την ανατροφοδότηση από πολλαπλές επιχειρήσεις χρηστών.

2.11 Μελέτη βελτίωσης της λήψης επιχειρηματικών αποφάσεων με βάση τους KPI από ένα σύστημα διαχείρισης

Η μελέτη [AuSa17] εισάγει μια προσέγγιση για τη βελτίωση της διαχείρισης των επιχειρήσεων σύμφωνα με τις αρχές της διαθεσιμότητας και ευελιξίας των πληροφοριών. Για την υποστήριξη αυτής της προσέγγισης, οι συγγραφείς ανέπτυξαν ένα φιλικό προς το χρήστη και αποτελεσματικό σύστημα που ονομάζεται BIM (σύστημα διαχείρισης επιχειρησιακών δεικτών) και μπορεί να χρησιμοποιηθεί από οποιαδήποτε επιχείρηση. Με βάση μια λεπτομερή μελέτη σε μια μεγάλης κλίμακας εταιρεία, αξιολόγησαν οι συγγραφείς την αποτελεσματικότητα του BIM σύμφωνα με τη χρηστικότητα του συστήματος, την ευελιξία των δεδομένων και την αποτελεσματικότητα της λήψης αποφάσεων. Το BIM ήταν επωφελής, δεδομένου ότι τα στελέχη μπορούσαν να συμβουλευούνται γρήγορα τους KPIs σε οποιαδήποτε στιγμή από οποιοδήποτε μέρος, αυξάνοντας έτσι τη χρονική αποδοτικότητα και ποιότητα του έργου λήψης αποφάσεων.

2.12 Boomerang: Προληπτικές συστάσεις με βάση τη διορατικότητα για την ανάλυση των δεδομένων συνομιλίας

Η διεπαφή σε φυσικές γλώσσες κερδίζει όλο και μεγαλύτερη δημοτικότητα λόγω της δυνατότητας να παρέχουν πρόσβαση σε δεδομένα και πληροφορίες, καθιστώντας την αλληλεπίδραση με τα δεδομένα πιο φυσική και προσιτή για ένα ευρύ φάσμα για τους χρήστες των επιχειρήσεων. Για την πλήρη υιοθέτηση του στόχου της επερώτησης μέσω φυσικής γλώσσας, είναι απαραίτητο να υπάρχει υποστήριξη καθοδήγησης για την εξερεύνηση των δεδομένων. Στο άρθρο [LQKÖ21] περιγράφεται το Boomerang, το οποίο είναι ένα σύστημα που προσφέρει την καθοδήγηση στην εξερεύνηση ενός συνόλου μέσω μιας διεπαφής συνομιλίας. Το Boomerang συγκεντρώνει προτάσεις από μια ποικιλία στατιστικών, συνεργατικών και βασισμένων στο περιεχόμενο των προτάσεων, και επιλέγει τις γνώσεις που ταιριάζουν στενά με τις ανάγκες του χρήστη την τρέχουσα κατάσταση της εξερεύνησης των δεδομένων, η οποία αντιπροσωπεύεται στο πλαίσιο συνομιλίας. Το Boomerang συνδυάζει διάφορες μετρήσεις, όπως τη συνάφεια, το ενδιαφέρον και την επικαιρότητα, για να κατατάξει τις ιδέες και να προτείνει τις γνώσεις με βάση το τρέχον πλαίσιο συνομιλίας. Επιπλέον, οι συγγραφείς δείχνουν πώς το σύστημα επιτρέπει την καθοδηγούμενη εξερεύνηση των δεδομένων σε ένα σύνολο πωλήσεων, που περιέχει πληροφορίες για προϊόντα, λιανοπωλητές, πωλήσεις, παραγγελίες, επίπεδα αποθεμάτων και περιοχές. Το σύνολο των δεδομένων περιέχει λεπτομέρειες σχετικά με τα μικτά περιθώρια κέρδους μέσω των πωλήσεων των διαφορετικών εμπορικών σημάτων προϊόντων σε διαφορετικές περιοχές από διαφορετικούς λιανοπωλητές.

2.13 Μελέτη MetaInsight: Αυτόματη ανακάλυψη δομημένης διερευνητικής γνώσης για διερευνητική ανάλυση δεδομένων

Η αυτόματη διερευνητική ανάλυση δεδομένων (EDA) επικεντρώνεται στην αυτόματη ανακάλυψη κομματιών γνώσης με τη μορφή μοτίβων. Ωστόσο, η γνώση που μεταφέρεται από αυτά τα προτεινόμενα δεδομένα είναι ασύνδετα ή στερούνται οργάνωση. Είναι δύσκολο για τους χρήστες να αποκτήσουν δομημένη γνώση. Επειδή ο αριθμός των προτεινόμενων μοτίβων αυξάνεται, αυτά τα μεμονωμένα μοτίβα είναι λιγότερο πιθανό να παρακινήσουν τους χρήστες να διεξάγουν ανάλυση, γεγονός που παρεμποδίζει τα προτεινόμενα μοτίβα να αξιοποιηθούν αποτελεσματικά για τη διευκόλυνση της EDA.

Στην μελέτη [MDHZ21], οι συγγραφείς προτείνουν μια δομημένη αναπαράσταση της γνώσης που εξάγεται από πολυδιάστατα δεδομένα, η οποία αποσκοπεί στην αποτελεσματική διευκόλυνση της EDA και ονομάζεται MetaInsight. Συγκεκριμένα, οι συγγραφείς προτείνουν μια νέα διατύπωση βασικών προτύπων για την καταγραφή των βασικών χαρακτηριστικών της κατανομής των δεδομένων, ώστε να επιτευχθεί η εξαγωγή της γνώσης. Στη συνέχεια, με βάση την εξόρυξη δεδομένων τα ομοιογενή πρότυπα (HDPs) και η ομοιότητα μεταξύ των προτύπων, προσδιορίζουν την MetaInsights με την κατηγοριοποίηση βασικών προτύπων (εντός ενός HDP) σε κοινά και εξαιρέσεις, επιτυγχάνοντας έτσι δομημένη αναπαράσταση γνώσης. Τα κοινά και οι εξαιρέσεις συγκεντρώνουν τη γνώση που λαμβάνεται από τις διαδικασίες επαγωγής και επικύρωσης, οι οποίες είναι δύο τυπικοί μηχανισμοί ανάλυσης που διεξάγονται στην EDA. Δίνεται μια νέα συνάρτηση βαθμολόγησης για την ποσοτικοποίηση της χρησιμότητας των MetaInsights, μιας αποτελεσματικής διαδικασίας εξόρυξης και έναν αλγόριθμο κατάταξης για την αυτόματη ανακάλυψη υψηλής ποιότητας MetaInsights από πολυδιάστατα δεδομένα. Αποδεικνύουν την αποτελεσματικότητα και την αποδοτικότητα των MetaInsight (όσον αφορά τη διευκόλυνση της EDA) μέσω πειραμάτων σε πραγματικές σύνολα δεδομένων και μελέτες χρηστών τόσο σε έμπειρους όσο και σε μη έμπειρους χρήστες.

ΚΕΦΑΛΑΙΟ 3

ΜΗΧΑΝΙΣΜΟΣ ΑΥΤΟΜΑΤΗΣ ΠΑΡΑΓΩΓΗΣ

ΑΝΑΦΟΡΩΝ

3.1 Παραγωγή KPI

3.2 Σύγκριση KPI με παρόμοιες τιμές

3.3 SQL Ερωτήσεις

3.4 Παραδείγματα

Στην παρούσα ενότητα περιγράφεται ο μηχανισμός που υλοποιήθηκε για την παραγωγή αυτόματων αναφορών.

3.1 Παραγωγή KPI

Τα δεδομένα, για τα οποία θα γίνει η μελέτη, θεωρούμε ότι είναι αποθηκευμένα σε έναν πολυδιάστατο πίνακα, δηλαδή σε έναν κύβο. Ένας KPI διαθέτει έναν κύριο πίνακα, ο οποίος συσχετίζεται με πίνακες διαστάσεων. Μπορεί στον KPI να οριστεί ένα σύνολο από φίλτρα για καθένα από τους πίνακες συσχέτισης, με βάση αυτά θα βρεθούν παρόμοιες τιμές, ώστε να γίνουν οι συγκρίσεις, να οριστεί ως μια συναθροιστική συνάρτηση με μία ή περισσότερες μετρήσεις (measures). Η μορφή ενός KPI στηρίζεται στην ανάκτηση και τον χαρακτηρισμό μιας συναθροιστικής τιμής η οποία προκύπτει από μια ερώτηση της μορφής:

```

SELECT aggregationFunction(Fact.measure)
FROM Fact
JOIN Dim1 ON Fact.Dim1_ID = Dim1.ID
JOIN Dim2 ON Fact.Dim2_ID = Dim2.ID
.
.
JOIN Dimk ON Fact.Dimk_ID = Dimk.ID
WHERE Dim1.A1 = value1 AND Dim2.A2 = value2 AND ..... AND Dimk.Ak = valuek

```

Σχήμα 3.1 Η γενική μορφή ενός KPI σε SQL

- Στο SELECT συναθροίζουμε measure από τον Fact Table
- Στο FROM καταγράφεται ο Fact Table
- Για κάθε διάσταση με φίλτρο (η παράμετρος και η τιμή που έχει δηλώσει ο χρήστης) δημιουργείται JOIN με τις συνθήκες συσχέτισης
- Στο WHERE δηλώνονται τα φίλτρα για τις διαστάσεις που έχει δηλώσει ο χρήστης

Για να καταχωρήσουμε ένα KPI πρέπει να προδιαγράψουμε:

1. τα διαπιστευτήρια για τη διασύνδεση στη βάση που θα γίνει η ανάλυση
2. τον κύριο πίνακα που θα γίνει η ανάλυση (Fact Table)
3. Μετρήσεις που θα γίνει η ανάλυση
4. Φίλτρα

Στάδια της ροής για την καταχώρηση KPI

- 1) Εισαγωγή στοιχείων της Βάσης Δεδομένων Ανάλυσης από τον χρήστη
- 2) Εμφάνιση όλων των διαθέσιμων οντοτήτων της Βάσης Δεδομένων
- 3) Καταχώρηση από τον χρήστη Fact Table
- 4) Εμφάνιση στον χρήστη όλοι οι πίνακες διαστάσεων του Fact Table
- 5) Καταχώρηση παραμέτρων από τον χρήστη στους πίνακες διαστάσεων και επίπεδο έρευνας
- 6) Καταχώρηση από τον χρήστη Φράγματα, Μετρήσεις
- 7) Αποθήκευση στοιχείων στην βάση δεδομένων

Για τη διαδικασία καταχώρησης ενός KPI υπάρχει παράδειγμα στην ενότητα 3.5.1. Παρακάτω για κάθε στάδιο της ροής καταχώρησης του KPI αναλύονται λεπτομερώς τα στοιχεία που το απαρτίζουν.

Στοιχεία Βάσης Δεδομένων Ανάλυσης:

Τα πεδία που καλείται να δηλώσει ο χρήστης, ώστε να γίνει η διασύνδεση με τη βάση δεδομένων ανάλυσης είναι τα εξής:

- **Name:** Όνομα του KPI
- **Sever:** Η IP της βάσης
- **Username:** Όνομα χρήστη για τη σύνδεση στη βάση
- **Password:** Κωδικός για τη σύνδεση στη βάση
- **Database:** Η βάση δεδομένων που θα χρησιμοποιήσουμε

Επιλογή Fact Table

Ο χρήστης καλείται να δηλώσει με ποια από τις οντότητες της βάσης δεδομένων θα γίνει η ανάλυση, το πεδίο είναι:

- **FactTable:** Ο κύριος πίνακας που θα χρησιμοποιηθεί για τον έλεγχο

Φράγματα

Τα πεδία που καλείται να δηλώσει ο χρήστης σαν Φράγματα για την ανάλυση του KPI είναι τα εξής:

- **Upthreshold:** Άνω φράγμα κατά τον έλεγχο του KPI
- **Downtreshold:** Κάτω φράγμα κατά τον έλεγχο του KPI
- **Goal:** Στόχος κατά τον έλεγχο του KPI

Μετρήσεις

Ο χρήστης καλείται να δηλώσει τα πεδία για τις πληροφορίες που επιθυμεί να δει στα αποτελέσματα κατά την εκτέλεση της ερώτησης στη βάση δεδομένων:

- **AttributeNameAtDB:** Όνομα παραμέτρου στη βάση
- **MeasureName:** Όνομα για την μέτρηση
- **Aggregation:** Παράμετρος που δεν είναι υποχρεωτική και ο χρήστης μπορεί να θέσει αν η τιμή επιθυμεί να είναι άθροισμα, μέσος όρος (sum, avg) κλπ.

Φίλτρα για τους πίνακες ανάλυσης

Αυτά τα πεδία τα ορίζει ο χρήστης εφόσον επιθυμεί περαιτέρω ανάλυση για την ερώτηση και μπορεί να δηλώσει τις εξής παραμέτρους:

- **AttributeName**: Αφορά την παράμετρο που θέλει να ορίσει ο χρήστης για το φιλτράρισμα
- **Value**: Η τιμή της παραμέτρου
- **IsTimeDimension**: Αν είναι Datetime η παράμετρος
- **AttributeMumName**: Αφορά την παράμετρο που θέλει να ορίσει ο χρήστης για το ανώτερο επίπεδο, από το οποίο θα βρεθούν οι παρόμοιες τιμές

Αποθήκευση δεδομένων

Κατά τη διαδικασία αποθήκευσης των δεδομένων, γίνεται έλεγχος αν σε κάποιο από τους πίνακες διαστάσεων έχει οριστεί παράμετρος και παράμετρος για το ανώτερο επίπεδο. Αν έχουν οριστεί αυτές οι παράμετροι, θα γίνει αποθήκευση και το ανώτερο επίπεδο, ώστε όταν γίνει η εκτέλεση του KPI με βάση αυτό θα βρεθούν οι παρόμοιες τιμές.

3.2 Σύγκριση KPI με παρόμοιες τιμές

3.2.1 Σύγκριση KPI με παρόμοιες τιμές, εκτέλεση ερώτησης για κάθε παρόμοια τιμή

Για την εκτέλεση ενός KPI θα δημιουργηθεί ένα πλήθος ερωτήσεων προς εκτέλεση στη βάση δεδομένων. Πιο αναλυτικά, αυτό σημαίνει, ότι για κάθε παράμετρο που δήλωσε ο χρήστης στους πίνακες διαστάσεων, θα βρεθούν οι παρόμοιες τιμές (αδέρφια τιμές) και για την καθεμία από αυτές θα δημιουργηθεί ερώτηση προς εκτέλεση στη βάση δεδομένων.

Θεωρούμε ότι έχουμε ένα KPI $I = \langle \text{agg}(M), \Phi \rangle$, όπου:

- agg η συναθροιστική συνάρτηση
- M το μέτρο στο οποίο γίνεται το συναθροιστικό μέτρο
- Φ μια λίστα ατομικών φίλτρων της μορφής $\text{Dimension}_i.L_i = u_i$

Η μετάφραση του KPI είναι όπως στο σχήμα 3.1.

Αλγόριθμος παραγωγής αδερφών ερωτήσεων

Αλγόριθμος 3.1 Παραγωγή αδερφών ερωτήσεων

-
- 1) Είσοδος: Ένα KPI
 - 2) Παράμετροι: topLevelValue = το string αποτέλεσμα του ανώτερου επιπέδου
 - 3) Για κάθε παράμετρο που έχει δηλώσει ο χρήστης στους πίνακες διαστάσεων
 - a) Γίνεται ανάκτηση του ανώτερου επιπέδου που δήλωσε ο χρήστης
 - b) topLevelValue = Εκτέλεση ερώτησης στη βάση, για την τιμή ανώτερου επιπέδου
 - c) Εκτέλεση ερώτησης στη βάση, για να βρεθούν οι παρόμοιες τιμές με φίλτρο την τιμή topLevelValue που βρέθηκε
 - d) Έξοδος: οι παρόμοιες τιμές (τιμές αδέρφια)
-

Για να γίνει η σύγκριση ενός KPI με τις παρόμοιες τιμές, θα πρέπει για την τιμή της παράμετρου που έχει δηλώσει ο χρήστης να βρεθούν οι παρόμοιες τιμές, δηλαδή τα αδέρφια του.

Για παράδειγμα, έστω ότι ο χρήστης έχει δηλώσει ένα KPI με 2 φίλτρα έστω (α) προς την παράμετρο city με τιμή Acapulco για τον πίνακα διάστασης store και (β) Product = Great Muffins για τον πίνακα διάστασης Product. Θα παραχθούν για κάθε φίλτρο αποτελέσματα ως προς τις παρόμοιες τιμές. Συγκεκριμένα:

Για τα «αδέρφια» του Acapulco, θα παραχθούν φίλτρα με τιμή Merida, Camacho, Mexico City, Guadalajara, Orizaba, Hidalgo, San Andres για την παράμετρο city και για καθεμία από αυτές θα γίνει ερώτηση στη βάση δεδομένων ώστε να βρεθεί το συνολικό ποσό της μέτρησης κρατώντας σταθερό το δεύτερο φίλτρο Product = Great Muffins.

Η ίδια διαδικασία θα γίνει και για το δεύτερο φίλτρο: θα βρεθούν οι τιμές «αδέρφια» του Product = Great Muffins οι οποίες στα δεδομένα της βάσης είναι Great Muffin, Great Bagels, Great Blueberry Muffins, Great Cranberry Muffins, Great English Muffins, Great Pumpernickel Bread, Great Rye Bread, Great Wheat Bread, Great White Bread και για καθεμία από αυτές θα γίνει ερώτηση στη βάση δεδομένων ώστε να βρεθεί το συνολικό ποσό της μέτρησης κρατώντας σταθερό το πρώτο φίλτρο city = Acapulco.

Οι τιμές αυτές για να εντοπιστούν μπορούν να συμβούν τα εξής 2 σενάρια:

- 1) Αν η παράμετρος που δηλώθηκε δεν είναι ξένο κλειδί σε άλλο πίνακα, αρχικά, θα δημιουργηθεί η ερώτηση για την εκτέλεση στη βάση δεδομένων, ώστε να βρεθεί η

τιμή του ανώτερου επιπέδου όταν ο πίνακας διάστασης έχει την τιμή που δήλωσε ο χρήστης και θα είναι της μορφής:

Αλγόριθμος 3.2 Εύρεση τιμής ανωτέρου επιπέδου

```
SELECT DISTINCT topVelel.attribute
FROM FactTable
JOIN lowLevel ON FactTable.DimensionTable_ID = lowLevel.ID
JOIN Dim2 ON Fact. Dim2_ID = Dim2.ID
.
.
JOIN Dimk ON Fact. Dimk_ID = Dimk.ID
Where lowLevel.Attribute = value AND Dim2.A2 = value2 AND ..... AND Dimk.Ak =
valuek
```

Για την τιμή που βρέθηκε, θα δημιουργηθεί η ερώτηση για την εκτέλεση στη βάση, ώστε να βρεθούν οι παρόμοιες τιμές για την ανάλυση του KPI και θα είναι της μορφής:

Αλγόριθμος 3.3 Εύρεση παρόμοιων τιμών

```
SELECT DISTINCT lowLevel.Attribute
FROM DimensionTable
Where topLevel.Attribute = topLevelValue
```

- 2) Αν η παράμετρος που δηλώθηκε είναι ξένο κλειδί σε άλλο πίνακα, αρχικά, θα δημιουργηθεί η ερώτηση για την εκτέλεση στη βάση δεδομένων, ώστε να βρεθεί η τιμή του ανώτερου επιπέδου όταν ο πίνακας διάστασης έχει την τιμή που δήλωσε ο χρήστης και θα γίνει η διασύνδεση (join) με τον πίνακα που υπάρχει συσχέτιση. Η μορφή της ερώτησης θα είναι:

Αλγόριθμος 3.4 Εύρεση τιμής ανωτέρου επιπέδου και συσχέτιση με άλλο πίνακα

```
SELECT topLevel.attribute
FROM lowLevel.DimensionTable
JOIN topLevel.TableName ON topLevel.private_key = lowLevel.foreign_key
JOIN Dim2 ON Fact. Dim2_ID = Dim2.ID
.
.
JOIN Dimk ON Fact. Dimk_ID = Dimk.ID
Where lowLevel.Attribute = value AND Dim2.A2 = value2 AND ..... AND Dimk.Ak =
valuek
```

Για την τιμή που βρέθηκε, θα δημιουργηθεί η ερώτηση για την εκτέλεση στη βάση, ως προς τον πίνακα διαστάσεων και της διασύνδεσης (join) με τον πίνακα συσχέτισης, ώστε να βρεθούν οι παρόμοιες τιμές για την ανάλυση του KPI και θα είναι της μορφής:

Αλγόριθμος 3.5 Εύρεση παρόμοιων τιμών και συσχέτιση με άλλο πίνακα

```
SELECT lowLevel.Attribute
FROM DimensionTable
JOIN topLevel.TableName ON topLevel.private_key = lowLevel.foreign_key
Where topLevel.Attribute = topLevelValue
```

Αλγόριθμος εκτέλεσης όλων των σταδίων ενός KPI

Αλγόριθμος 3.6 Εκτέλεση όλων των σταδίων ενός KPI

-
- 1) Είσοδος: Η προδιαγραφή ενός KPI με τη βάση, Fact Table, Διαστάσεις, φίλτρα και μετρήσεις στην κατάλληλη μορφή
 - 2) Παράμετρος: query = το string της sql ερώτησης που θα εκτελεστεί χωρίς την τιμή της παραμέτρου που γίνεται η ανάλυση
 - 3) Παράμετρος: attributeValue = η τιμή της παραμέτρου
 - 4) Για κάθε πίνακα διαστάσεων για τον οποίο έχει δηλώσει παράμετρο ο χρήστης
 - a) Εκτέλεση στη βάση ερώτηση για τις παρόμοιες τιμές (αδέρφια)
 - b) Για κάθε αποτέλεσμα των παρόμοιων τιμών
 - i) Προσθήκη στην παράμετρο attributeValue την παρόμοια τιμή
 - ii) Κατασκευή νέας ερώτησης και εισαγωγή στην παράμετρο query
 - iii) Προσθήκη στην ερώτηση (query) την παράμετρο attributeValue
 - iv) Προσθήκη στην ερώτηση (query) τα υπόλοιπα φίλτρα των διαστάσεων
 - v) Εκτέλεση της ερώτησης (query)
 - 5) Έξοδος: Όλα τα αποτελέσματα των εκτελέσεων
-

Εύρεση όλων των Μετρήσεων και καταγραφή στην κατάλληλη μορφή

Με βάση το KPI που επιλέχθηκε από τον χρήστη γίνεται ανάκτηση των δεδομένων των μετρήσεων από τη βάση. Αν ο χρήστης έχει δηλώσει στο πεδίο aggregation τιμή τότε η μέτρηση θα έχει την μορφή aggregation(τιμή_μέτρησης), όπου aggregation sum, avg κλπ.

Εκτέλεση KPI

Η ερώτηση που δημιουργείται θα εκτελεστεί στη βάση δεδομένων για κάθε πίνακα διαστάσεων για τον οποίο έχει δηλώσει ο χρήστης παράμετρο και για όσες παρόμοιες τιμές (αδέρφια) βρεθούν για την τιμή αυτή. Για κάθε εκτέλεση θα προστεθούν και τα υπόλοιπα φίλτρα, δηλαδή τις παραμέτρους των υπόλοιπων διαστάσεων που δήλωσε ο χρήστης.

3.2.2 Σύγκριση KPI με παρόμοιες τιμές, εκτέλεση μιας ερώτησης με βάση την ανώτερη τιμή

Για την εκτέλεση ενός KPI θα δημιουργηθεί μια ερώτηση προς εκτέλεση στη βάση δεδομένων. Πιο αναλυτικά, αυτό σημαίνει, ότι για κάθε παράμετρο που δήλωσε ο χρήστης στους πίνακες διαστάσεων, θα βρεθεί το αποτέλεσμα του ανώτερου επιπέδου ως προς την τιμή που δήλωσε ο χρήστης στον πίνακα διαστάσεων, και θα γίνει ομαδοποίηση για την παράμετρο αυτή. Με την εκτέλεση της ερώτησης στη βάση δεδομένων θα βρεθούν οι παρόμοιες τιμές (αδέρφια τιμές).

Αλγόριθμος παραγωγής ερώτησης

Αλγόριθμος 3.7 Παραγωγή ερώτησης για παρόμοιες τιμές και μετρήσεις

-
- 4) Είσοδος: Ένα KPI
 - 5) Παράμετροι: `topLevelValue` = το string αποτέλεσμα του ανώτερου επιπέδου
 - 6) Για κάθε παράμετρο που έχει δηλώσει ο χρήστης στους πίνακες διαστάσεων
 - a) Γίνεται ανάκτηση του ανώτερου επιπέδου που δήλωσε ο χρήστης
 - b) `topLevelValue` = Εκτέλεση ερώτησης στη βάση, για την τιμή ανώτερου επιπέδου
 - c) Εκτέλεση ερώτησης στη βάση, για να βρεθούν οι παρόμοιες τιμές και τα ποσά της μέτρησης με φίλτρο την τιμή `topLevelValue` που βρέθηκε
 - d) Έξοδος: οι παρόμοιες τιμές (τιμές αδέρφια) και τα ποσά της μέτρησης
-

Για να γίνει η σύγκριση ενός KPI με τις παρόμοιες τιμές, θα πρέπει για την τιμή της παράμετρου που έχει δηλώσει ο χρήστης να βρεθούν οι παρόμοιες τιμές, δηλαδή τα αδέρφια του. Οι τιμές αυτές για να εντοπιστούν μπορούν να συμβούν τα εξής 2 σενάρια:

- 1) Αν η παράμετρος που δηλώθηκε δεν είναι ξένο κλειδί σε άλλο πίνακα, αρχικά, θα δημιουργηθεί η ερώτηση για την εκτέλεση στη βάση δεδομένων, ώστε να βρεθεί η τιμή του ανώτερου επιπέδου όταν ο πίνακας διάστασης έχει την τιμή που δήλωσε ο χρήστης και θα είναι της μορφής:

Αλγόριθμος 3.8 Εύρεση τιμής ανωτέρου επιπέδου

```
SELECT topVelel.attribute  
FROM FactTable  
JOIN lowLevel ON FactTable.DimensionTable_ID = lowLevel.ID  
JOIN Dim2 ON Fact. Dim2 _ID = Dim2.ID  
.  
.  
JOIN Dimk ON Fact. Dimk _ID = Dimk.ID  
Where lowLevel.Attribute = value AND Dim2.A2 = value2 AND ..... AND Dimk.Ak =  
valuek
```

Για την τιμή που βρέθηκε, θα δημιουργηθεί η ερώτηση για την εκτέλεση στη βάση, ώστε να βρεθούν τα αποτελέσματα με τις παρόμοιες τιμές για την ανάλυση του KPI και θα είναι της μορφής:

Αλγόριθμος 3.9 Αποτελέσματα παρόμοιων τιμών με τις μετρήσεις

```
SELECT lowLevel.attribute, sum(measure)  
FROM FactTable  
JOIN DimensionTable ON FactTable.DimensionTable_ID = DimensionTable.ID  
Where DimensionTable .topVelelAttribute = highLevelvalue  
GROUP BY lowLevel.attribute
```

- 2) Αν η παράμετρος που δηλώθηκε είναι ξένο κλειδί σε άλλο πίνακα, αρχικά, θα δημιουργηθεί η ερώτηση για την εκτέλεση στη βάση δεδομένων, ώστε να βρεθεί η τιμή του ανώτερου επιπέδου όταν ο πίνακας διάστασης έχει την τιμή που δήλωσε ο

χρήστης και θα γίνει η διασύνδεση (join) με τον πίνακα που υπάρχει συσχέτιση. Η μορφή της ερώτησης θα είναι:

Αλγόριθμος 3.10 Εύρεση τιμής ανωτέρου επιπέδου με συσχέτιση σε άλλο πίνακα

```
SELECT topVelel.attribute
FROM lowLevel.DimensionTable
JOIN topLevel.TableName ON topLevel.private_key = lowLevel.foreign_key
JOIN Dim2 ON Fact. Dim2 _ID = Dim2.ID
.
.
JOIN Dimk ON Fact. Dimk _ID = Dimk.ID
Where lowLevel.Attribute = value AND Dim2.A2 = value2 AND ..... AND Dimk.Ak =
valuek
```

Για την τιμή που βρέθηκε, θα δημιουργηθεί η ερώτηση για την εκτέλεση στη βάση, ως προς τον πίνακα διαστάσεων και της διασύνδεσης (join) με τον πίνακα συσχέτισης, ώστε να βρεθούν τα αποτελέσματα για τις παρόμοιες τιμές για την ανάλυση του KPI και θα είναι της μορφής:

Αλγόριθμος 3.11 Αποτελέσματα παρόμοιων τιμών με τις μετρήσεις

```
SELECT lowLevel.attribute, sum(measure)
FROM FactTable
JOIN DimensionTable ON FactTable.DimensionTable_ID = DimensionTable.ID
JOIN topLevel.TableName ON topLevel.private_key = lowLevel.foreign_key
Where DimensionTable .topVelelAttribute = highLevelvalue
Group by lowLevel.attribute
```

Αλγόριθμος εκτέλεσης όλων των σταδίων ενός KPI

Αλγόριθμος 3.12 Εκτέλεση όλων των σταδίων ενός KPI

-
- 1) **Είσοδος:** Η προδιαγραφή ενός KPI με τη βάση, Fact Table, Διαστάσεις, φίλτρα και μετρήσεις στην κατάλληλη μορφή
 - 2) Παράμετρος: query = το string της sql ερώτησης που θα εκτελεστεί χωρίς την τιμή της παραμέτρου του ανώτερου επιπέδου που γίνεται η ανάλυση
 - 3) Παράμετρος: highLevelValue = η τιμή της παραμέτρου του ανώτερου επιπέδου
 - 4) Για κάθε πίνακα διαστάσεων για τον οποίο έχει δηλώσει παράμετρο ο χρήστης
 - a) highLevelValue = Εκτέλεση στη βάση ερώτηση για την εύρεση τιμής του ανωτέρου επιπέδου
 - b) Προσθήκη στην ερώτηση (query) την παράμετρο highLevel και την τιμή που βρέθηκε highLevelValue
 - c) **Έξοδος:** Το αποτελέσματα της εκτέλεσης
-

Εύρεση όλων των Μετρήσεων και καταγραφή στην κατάλληλη μορφή

Με βάση το KPI που επιλέχθηκε από τον χρήστη γίνεται ανάκτηση των δεδομένων των μετρήσεων από τη βάση. Αν ο χρήστης έχει δηλώσει στο πεδίο aggregation τιμή τότε η μέτρηση θα έχει την μορφή aggregation(τιμή_μέτρησης), όπου aggregation sum, avg κλπ.

Εκτέλεση KPI

Η ερώτηση που δημιουργείται θα εκτελεστεί στη βάση δεδομένων για κάθε πίνακα διαστάσεων για τον οποίο έχει δηλώσει ο χρήστης παράμετρο. Για κάθε εκτέλεση θα προστεθούν και τα υπόλοιπα φίλτρα, δηλαδή τις παραμέτρους των υπόλοιπων διαστάσεων που δήλωσε ο χρήστης.

3.3 Σύγκριση της πραγματικής τιμής του KPI με την αναμενόμενη τιμή ως προς το ανώτερο επίπεδο ή ως προς το σύνολο

Ένας άλλος τρόπος να αποτιμήσουμε ένα KPI είναι να προσπαθήσουμε να συγκρίνουμε τον KPI με μια αναμενόμενη τιμή για αυτόν.

Θεωρούμε ότι έχουμε ένα KPI $I = \langle \text{agg}(M), \Phi \rangle$, όπου:

- agg η συναθροιστική συνάρτηση
- M το μέτρο στο οποίο γίνεται το συναθροιστικό μέτρο

— Φ μια λίστα ατομικών φίλτρων της μορφής Dimension_i.Li = ui
 Η μετάφραση του KPI είναι όπως στο σχήμα 3.1.

Ο πιο απλός τρόπος αφορά την περίπτωση που προσδιορίζεται ένα χρονικό φίλτρο (π.χ., οι πωλήσεις κρασιών στην Ήπειρο το 2019). Σε αυτή την περίπτωση μια απλή σύγκριση είναι με τις πωλήσεις σε σχέση με την προηγούμενη χρονιά. Πόσο όμως παραπάνω ή παρακάτω πρέπει να είναι η τιμή του KPI σε σχέση με πέρυσι ? Υπάρχουν δύο πιθανές απαντήσεις:

- 1) σε σχέση με το πως πήγαν οι πωλήσεις ως προς το ανώτερο επίπεδο που ανήκει το φίλτρο σε σχέση με την προηγούμενη χρονιά (ενότητα 3.3.1), οπότε η αναμενόμενη τιμή είναι:

$$\text{περσινή_τιμή_KPI} * \frac{\text{συνολική_ανωτέρουΕπιπέδου_φετινή_τιμή}}{\text{συνολική_ανωτέρουΕπιπέδου_περσινή_τιμή}}$$

- 2) σε σχέση με το πως πήγαν συνολικά οι πωλήσεις σε σχέση με την προηγούμενη χρονιά (ενότητα 3.3.2), οπότε η αναμενόμενη τιμή είναι:

$$\text{περσινή_τιμή_KPI} * \frac{\text{συνολική_φετινή_τιμή}}{\text{συνολική_περσινή_τιμή}}$$

Για την σύγκριση της πραγματικής τιμής του KPI με την αναμενόμενη τιμή ως προς το ανώτερο επίπεδο και ως προς το σύνολο, θα πρέπει κατά την καταχώρηση των δεδομένων του KPI μαζί με τις παραμέτρους των διαστάσεων να έχει δηλώσει ο χρήστης και μια παράμετρο για το έτος για την οποία θα γίνει ανάλυση. Υπάρχει σχετικό παράδειγμα στην ενότητα 3.5.5.

3.3.1 Σύγκριση της πραγματικής τιμής του KPI με την αναμενόμενη τιμή ως προς το ανώτερο επίπεδο

Αλγόριθμος 3.13 Σύγκριση της πραγματικής τιμής του KPI με την αναμενόμενη τιμή ως προς το ανώτερο επίπεδο

-
- 1) **Είσοδος:** Ένα KPI
 - 2) **Παράμετροι:**
 - a) `topLevelValue` = μια μεταβλητή που κρατά τη αντίστοιχη τιμή του KPI για ένα επίπεδο πιο πάνω,
 - b) `dateValue` = η τιμή του KPI στην ημερομηνία για την ανάλυση,
 - c) `highLevelTotalThisYear` = το συνολικό ποσό της μέτρησης για το ανώτερο επίπεδο με την τιμή ημερομηνίας,
 - d) `highLevelTotalLastYear` = το συνολικό ποσό της μέτρησης για το ανώτερο επίπεδο για το προηγούμενο έτος της ημερομηνίας,
 - e) `actualValueThisYear` = το συνολικό ποσό της μέτρησης για την παράμετρο της διάστασης και με την τιμή ημερομηνίας,
 - f) `actualValueLastYear` = το συνολικό ποσό της μέτρησης για την παράμετρο της διάστασης και με την τιμή ημερομηνίας για το προηγούμενο έτος
 - 3) Για κάθε παράμετρο που έχει δηλώσει ο χρήστης στους πίνακες διαστάσεων
 - a) Γίνεται ανάκτηση του ανώτερου επιπέδου που δήλωσε ο χρήστης
 - b) `dateValue` = Γίνεται ανάκτηση της τιμής του χρόνου που θα γίνει η ανάλυση
 - c) `topLevelValue` = Εκτέλεση ερώτησης στη βάση, για την τιμή ανώτερου επιπέδου
 - d) `highLevelTotalThisYear` = Εκτέλεση ερώτησης στη βάση, για να βρεθεί το συνολικό ποσό της μέτρησης με φίλτρο την τιμή `topLevelValue` που βρέθηκε και για το έτος της `dateValue`
 - e) `highLevelTotalLastYear` = Εκτέλεση ερώτησης στη βάση, για να βρεθεί το συνολικό ποσό της μέτρησης με φίλτρο την τιμή `topLevelValue` που βρέθηκε και για το προηγούμενο έτος του `dateValue`
 - f) `actualValueThisYear` = Εκτέλεση ερώτησης στη βάση, για να βρεθεί το πραγματικό ποσό της μέτρησης με φίλτρο τη διάσταση και φίλτρο για το έτος της `dateValue`
 - g) `actualValueLastYear` = Εκτέλεση ερώτησης στη βάση, για να βρεθεί το πραγματικό ποσό της μέτρησης με φίλτρο τη διάσταση και φίλτρο για το προηγούμενο έτος της `dateValue`
 - h) **Έξοδος:** η αναμενόμενη τιμή
$$\text{actualValueLastYear} * (\text{highLevelTotalThisYear} / \text{highLevelTotalLastYear})$$
-

Για την υλοποίηση του αλγορίθμου σύγκρισης της πραγματικής τιμής του KPI με την αναμενόμενη τιμή ως προς το ανώτερο επίπεδο η διαδικασία θα γίνει για κάθε φίλτρο στους πίνακες διαστάσεων. Αρχικά θα δημιουργηθεί η ερώτηση για την εκτέλεση στη βάση δεδομένων, ώστε να βρεθεί η τιμή του ανώτερου επιπέδου, όταν ο πίνακας διάστασης έχει την τιμή που δήλωσε ο χρήστης και θα είναι της μορφής:

Αλγόριθμος 3.14 Εύρεση τιμής ανωτέρου επιπέδου

```
SELECT topVelel.attribute
FROM FactTable
  JOIN Dim2 ON Fact. Dim2 _ID = Dim2.ID
  .
  .
  JOIN Dimk ON Fact. Dimk _ID = Dimk.ID
Where lowLevel.Attribute = value AND Dim2.A2 = value2 AND ..... AND Dimk.Ak =
valuek
```

Για την τιμή που βρέθηκε, θα δημιουργηθεί η ερώτηση για την εκτέλεση στη βάση, ώστε να βρεθεί το συνολικό ποσό για το έτος που έχει δηλώσει ο χρήστης για την ανάλυση του KPI και θα είναι της μορφής:

Αλγόριθμος 3.15 Υπολογισμός για το συνολικό ποσό για το έτος τους ανωτέρου επιπέδου

```
SELECT sum(measure)
FROM FactTable
JOIN DimensionTable ON FactTable.DimensionTable_ID = DimensionTable.ID
JOIN   DimensionDateTable   ON   FactTable.DimensionDateTable_ID   =
DimensionDateTable.ID
Where topLevel.Attribute = topLevelValue and DimensionDateTable.Year =
Year
```

Η ίδια ερώτηση θα εκτελεστεί και για το προηγούμενο έτος. Η διαφορά στην ερώτηση θα είναι στην παράμετρο Year, όπου θα είναι Year -1.

Στη συνέχεια θα βρεθεί και το συνολικό ποσό για την παράμετρο που δήλωσε ο χρήστης στον πίνακα διαστάσεων και για την ημερομηνία που γίνεται η ανάλυση και θα είναι της μορφής:

Αλγόριθμος 3.16 Υπολογισμός για το συνολικό ποσό για το έτος της παραμέτρου

```
SELECT sum(measure)
FROM FactTable
JOIN DimensionTable ON FactTable.DimensionTable_ID = DimensionTable.ID
JOIN    DimensionDateTable    ON    FactTable.DimensionDateTable_ID    =
DimensionDateTable.ID
Where lowLevel.Attribute = lowLevelValue and DimensionDateTable.Year =
Year
```

Η ίδια ερώτηση θα εκτελεστεί και για το προηγούμενο έτος με τη διαφορά ότι η παράμετρος YearValue θα είναι Year – 1, ώστε να γίνει η σύγκριση με το προηγούμενο έτος.

Τέλος, θα υπολογιστεί η αναμενόμενη τιμή με βάση τα αποτελέσματα από τις εκτελέσεις των παραπάνω ερωτήσεων.

Αλγόριθμος 3.17 Τύπος αναμενόμενης τιμής

```
expectedthisYearValue =
actualLastYearValue * (highLevelTotalThisYear/ highLevelTotalLastYear )
```

3.3.2 Σύγκριση της πραγματικής τιμής του KPI με την αναμενόμενη τιμή ως προς το σύνολο

Αλγόριθμος υλοποίησης

Αλγόριθμος 3.18 Σύγκριση της πραγματικής τιμής του KPI με την αναμενόμενη τιμή ως προς το σύνολο

-
- 1) Είσοδος: Ένα KPI
 - 2) Παράμετροι:
 - a) `dateValue` = η τιμή της ημερομηνίας για την ανάλυση,
 - b) `totalThisYear` = το συνολικό ποσό της μέτρησης για την τιμή ημερομηνίας,
 - c) `totalLastYear` = το συνολικό ποσό της μέτρησης για το προηγούμενο έτος της ημερομηνίας,
 - d) `actualValueThisYear` = το συνολικό ποσό της μέτρησης για την παράμετρο της διάστασης και την τιμή ημερομηνίας,
 - e) `actualValueLastYear` = το συνολικό ποσό της μέτρησης για την παράμετρο της διάστασης και την τιμή ημερομηνίας για το προηγούμενο έτος
 - 3) Για κάθε παράμετρο που έχει δηλώσει ο χρήστης στους πίνακες διαστάσεων
 - a) Γίνεται ανάκτηση του ανώτερου επίπεδου που δήλωσε ο χρήστης
 - b) `dateValue` = Γίνεται ανάκτηση της τιμής του χρόνου που θα γίνει η ανάλυση
 - c) `totalThisYear` = Εκτέλεση ερώτησης στη βάση, για να βρεθεί το συνολικό ποσό της μέτρησης για το έτος της `dateValue`
 - d) `totalLastYear` = Εκτέλεση ερώτησης στη βάση, για να βρεθεί το συνολικό ποσό της μέτρησης για το προηγούμενο έτος του `dateValue`
 - e) `actualValueThisYear` = Εκτέλεση ερώτησης στη βάση, για να βρεθεί το πραγματικό ποσό της μέτρησης με φίλτρο τη διάσταση και φίλτρο για το έτος της `dateValue`
 - f) `actualValueLastYear` = Εκτέλεση ερώτησης στη βάση, για να βρεθεί το πραγματικό ποσό της μέτρησης με φίλτρο τη διάσταση και φίλτρο για το προηγούμενο έτος της `dateValue`
 - g) Έξοδος: η αναμενόμενη τιμή
$$\text{actualValueLastYear} * (\text{totalThisYear} / \text{totalLastYear})$$
-

Για να εκτελεστεί ο Αλγόριθμος σύγκρισης της πραγματικής τιμής του KPI με την αναμενόμενη τιμή ως προς το σύνολο, θα δημιουργηθεί η ερώτηση για την εκτέλεση στη βάση, ώστε να βρεθεί το συνολικό ποσό για το έτος που έχει δηλώσει ο χρήστης για την ανάλυση του KPI και θα είναι της μορφής:

Αλγόριθμος 3.19 Υπολογισμός για το συνολικό ποσό για το έτος

```
SELECT sum(measure)
FROM FactTable
JOIN    DimensionDateTable    ON    FactTable.DimensionDateTable_ID    =
DimensionDateTable.ID
Where DimensionDateTable.Year = Year
```

Η ίδια ερώτηση θα εκτελεστεί και για το προηγούμενο έτος. Η διαφορά στην ερώτηση θα είναι στην παράμετρο Year, όπου θα είναι Year -1.

Στη συνέχεια θα βρεθεί και το συνολικό ποσό για την παράμετρο που δήλωσε ο χρήστης στον πίνακα διαστάσεων και για την ημερομηνία που γίνεται η ανάλυση και θα είναι της μορφής:

Αλγόριθμος 3.20 Υπολογισμός για το συνολικό ποσό για το έτος

```
SELECT sum(measure)
FROM FactTable
JOIN DimensionTable ON FactTable.DimensionTable_ID = DimensionTable.ID
JOIN    DimensionDateTable    ON    FactTable.DimensionDateTable_ID    =
DimensionDateTable.ID
Where lowLevel.Attribute = lowLevelValue and DimensionDateTable.Year =
YearValue
```

Η ίδια ερώτηση θα εκτελεστεί και για το προηγούμενο έτος με τη διαφορά ότι η παράμετρος YearValue θα είναι YearValue - 1, ώστε να γίνει η σύγκριση με το προηγούμενο έτος.

Τέλος, θα υπολογιστεί η αναμενόμενη τιμή με βάση τα αποτελέσματα από τις εκτελέσεις των παραπάνω ερωτήσεων.

Αλγόριθμος 3.21 Τύπος αναμενόμενης τιμής για το σύνολο

```
expectedthisYearValue =
actualLastYearValue * (totalThisYear/ totalLastYear )
```

3.4 SQL Ερωτήσεις

Στην υποενότητα αυτή υπάρχουν οι ερωτήσεις και ο κώδικας παραγωγής ερωτήσεων για τις ανάγκες της διαδικασίας κατά την καταχώρηση ή την εκτέλεση KPI.

3.4.1 Ερώτηση Συσχέτισης πινάκων διαστάσεων με κάποιον πίνακα

Αλγόριθμος 3.22 Ερώτηση εύρεσης πινάκων συσχέτισης

```
SELECT
    sch.name AS schema_name,
    main_table.name AS main_table,
    main_column.name AS main_column,
    reference_table.name AS referenced_table,
    reference_column.name AS referenced_column
FROM sys.foreign_key_columns foreign_key_table
    INNER JOIN sys.objects system_objects ON system_objects.object_id =
foreign_key_table.constraint_object_id
    INNER JOIN sys.tables main_table ON main_table.object_id =
foreign_key_table.parent_object_id
    INNER JOIN sys.schemas sch ON main_table.schema_id = sch.schema_id
    INNER JOIN sys.columns main_column ON main_column.column_id =
parent_column_id AND main_column.object_id = main_table.object_id
    INNER JOIN sys.tables reference_table ON reference_table.object_id =
foreign_key_table.referenced_object_id
    INNER JOIN sys.columns reference_column ON reference_column.column_id =
referenced_column_id
    where reference_column.object_id = reference_table.object_id and main_table.name =
@table;
```

Η ερώτηση αυτή χρησιμοποιείται προκειμένου να βρεθούν οι πίνακες με τους οποίους συσχετίζεται ένας συγκεκριμένος πίνακας. Δίνεται το όνομα του στην παράμετρο @table και σαν αποτέλεσμα βλέπουμε:

- **schema_name**: το όνομα του σχήματος της βάσης
- **main_table**: το όνομα του πίνακα που έγινε η ερώτηση
- **main_column**: το ξένο κλειδί του πίνακα που έγινε η ερώτηση με τον πίνακα που συσχετίζεται
- **referenced_table**: το όνομα του πίνακα συσχέτισης
- **referenced_column**: το όνομα του private_key στον πίνακα συσχέτισης

Η συγκεκριμένη ερώτηση χρησιμοποιείται κατά την καταχώρηση ενός KPI. Όταν ο χρήστης δηλώσει πίνακα Fact Table γίνεται η ερώτηση στη βάση δεδομένων, ώστε να εμφανιστούν στον χρήστη οι πίνακες διαστάσεων που συσχετίζεται.

3.4.2 Ερώτηση για εύρεση τιμής του ανώτερου επιπέδου

Αλγόριθμος 3.23 Κώδικας παραγωγής της ερώτησης για τα αποτελέσματα-πίνακα συσχέτισης

```
string query = "SELECT DISTINCT " + bottomLevel.LevelMum.DimensionTable.Name +  
"." + bottomLevel.LevelMum.AttributeName + " as datavalue "  
    + " from " + factTable;  
query += joinTables;  
query += " where " + bottomLevel.DimensionTable.Name + "." +  
bottomLevel.AttributeName + " = @attributeValue ";  
if (!String.IsNullOrEmpty(whereClause))  
{  
    query += " and " + whereClause;  
}
```

Ο συγκεκριμένος κώδικας παραγωγής της ερώτησης χρησιμοποιείται για την εύρεση της τιμής του ανώτερου επιπέδου, για την τιμή που έχει το κατώτερο επίπεδο με βάση ό,τι έχει καταχωρήσει ο χρήστης. Αυτό σημαίνει ότι θα βρεθεί η τιμή του ανώτερου επιπέδου όταν το κατώτερο επίπεδο έχει την τιμή που έδωσε ο χρήστης στον πίνακα διαστάσεων για μια συγκεκριμένη παράμετρο. Η διαδικασία αυτή είναι μέρος της εύρεσης παρόμοιων τιμών.

3.4.3 Ερώτηση για αποτελεσμάτων παρόμοιων τιμών (αδερφών)

Αλγόριθμος 3.24 Κώδικας παραγωγής της ερώτησης για τα αποτελέσματα-αδέρφια

```
string query = "SELECT DISTINCT"
+ bottomLevel.DimensionTable.Name + "." + bottomLevel.AttributeName +
" as datavalue "
+ " from " + bottomLevel.DimensionTable.Name;

if(!highLevel.DimensionTable.Name.Equals(levelTocheck.DimensionTable.Name))
{
    query += " join " + highLevel.DimensionTable.Name + " on "
        + highLevel.DimensionTable.ConstraintCondition;
}
query += " where " + highLevel.DimensionTable.Name + "." + highLevel.AttributeName + " =
@attributeValue";
```

Ο συγκεκριμένος κώδικας παραγωγής της ερώτησης χρησιμοποιείται για την εύρεση των παρόμοιων τιμών, με βάση την τιμή που βρέθηκε από το ανώτερο επίπεδο. Τα αποτελέσματα αφορούν όλες τις τιμές αδέρφια για την τιμή που καταχώρησε ο χρήστης στην παράμετρο για κάποιον πίνακα διαστάσεων.

3.5 Παραδείγματα

3.5.1 Καταχώρηση KPI

Η διαδικασία καταχώρησης ενός KPI αφορά μια φόρμα, στην οποία ο χρήστης εισάγει τα δεδομένα στη διεπαφή της εφαρμογής. Αρχικά, ο χρήστης καλείται να καταχωρήσει τα στοιχεία της βάσης δεδομένων για την οποία θα γίνει η ανάλυση.

Name	
Server	
Username	
Password	
DataBase	

Σχήμα 3.2 Πεδία για τη βάση δεδομένων

Μόλις δηλώσει τα στοιχεία της βάσης πατάει το κουμπί Get Tables και εμφανίζονται στη διεπαφή της εφαρμογής όλοι οι διαθέσιμοι πίνακες καθώς και τα πεδία που αφορούν φράγματα (άνω, κάτω και στόχος) στα οποία ο χρήστης καλείται να δώσει τιμές.

FactTable	sales_fact_1998
Upthreshold	
Downthreshold	
Goal	

Σχήμα 3.3 Πεδίο Fact Table και φράγματα

Για να προσθέσει μετρήσεις για τον πίνακα Fact Table θα δηλώσει τα εξής πεδία καθώς και aggregation (sum, avg κλπ) για τη μέτρηση.

AttributeNameAtDB	
MeasureName	
Aggregation	

Σχήμα 3.4 Πεδία για τη Μέτρηση

Μόλις ο χρήστης επιλέξει Fact Table εμφανίζονται απευθείας όλοι οι πίνακες διαστάσεων με τους οποίους συσχετίζεται.

Name	store
ConstraintCondition	sales_fact_1998.store_id = store.store_id
IsTimeDimension	
Name	store
AttributeName	
Value	
AttributeMum	

Add Attribute

Name	customer
ConstraintCondition	sales_fact_1998.customer_id = customer.customer_id
IsTimeDimension	
Name	customer
AttributeName	
Value	
AttributeMum	

Add Attribute

Name	time_by_day
ConstraintCondition	sales_fact_1998.time_id = time_by_day.time_id
IsTimeDimension	
Name	time_by_day

Σχήμα 3.5 Πίνακες διαστάσεων με τις παραμέτρους σαν φίλτρα

Στα πλαίσια για κάθε πίνακα διαστάσεων που εμφανίστηκε στη διεπαφή της εφαρμογής μπορεί ο χρήστης να δηλώσει την παράμετρο, την τιμή καθώς και την τιμή της παραμέτρου του ανώτερου επιπέδου. Αν ο χρήστης επιθυμεί να προσθέσει παραπάνω από ένα φίλτρο στον ίδιο πίνακα διαστάσεων θα πατήσει το κουμπί Add Attribute.

3.5.2 Εύρεση πινάκων συσχέτισης με τον κύριο πίνακα (Fact Table)

Κατά τη δημιουργία ενός KPI μόλις δηλώσει ο χρήστης τον κύριο πίνακα (Fact Table) η διεπαφή αποστέλλει κλήση προς το Service, το οποίο συνδέεται με τη βάση δεδομένων

προκειμένου να πάρει ως αποτελέσματα τους πίνακες διαστάσεων, με τους οποίους συσχετίζεται.

Για το τρέχον παράδειγμα χρησιμοποιήθηκε η βάση δεδομένων northwind. Ο χρήστης δηλώνει τον πίνακα Order_Details και οι πίνακες που συσχετίζεται είναι ο Products και Orders.

The screenshot displays a web-based interface for configuring a Fact Table. At the top, a dropdown menu labeled 'FactTable' is set to 'Order_Details'. Below this are three input fields: 'Upthreshold', 'Downthreshold', and 'Goal'. A button labeled 'Add a new Measure' is positioned below these fields. A dashed line separates the Fact Table configuration from the dimension table configuration. Below the dashed line, there are two dimension table configurations. The first configuration is for a table named 'Products', with a 'ConstraintCondition' of 'Order_Details.ProductID = Products.Proc'. The second configuration is for a table named 'Orders', with a 'ConstraintCondition' of 'Order_Details.OrderID = Orders.OrderID'. Both dimension table configurations have an 'IsTimeDimension' field.

FactTable	Order_Details
Upthreshold	
Downthreshold	
Goal	

Add a new Measure

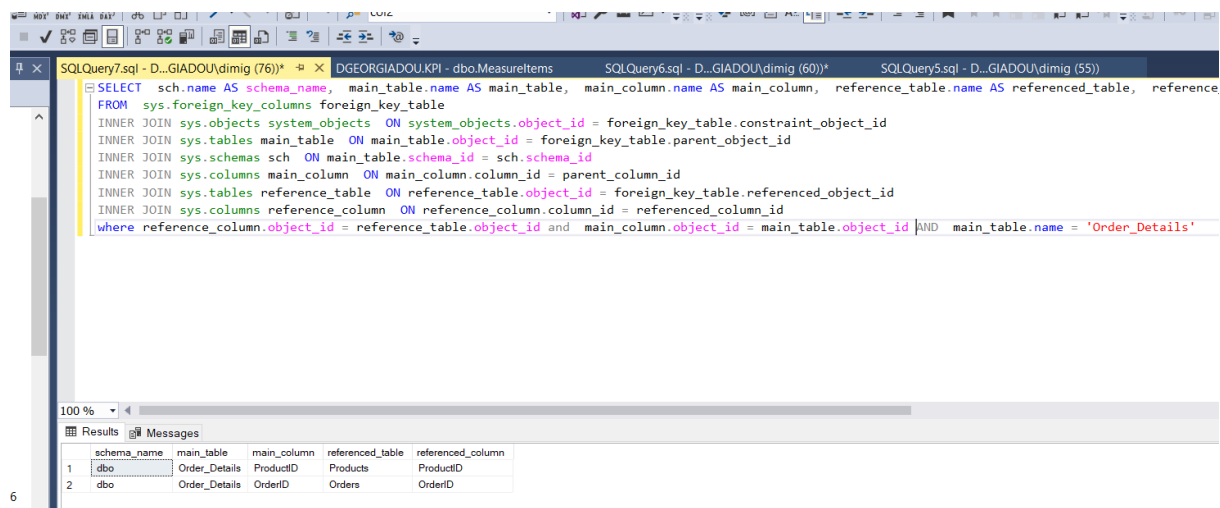
Name	Products
ConstraintCondition	Order_Details.ProductID = Products.Proc
IsTimeDimension	

Name	Products
AttributeName	
Value	

Name	Orders
ConstraintCondition	Order_Details.OrderID = Orders.OrderID
IsTimeDimension	

Σχήμα 3.6 UI Fact Table εμφάνιση Constraint Tables

Το αντίστοιχο παράδειγμα κατά την εκτέλεση της ερώτησης στη βάση δεδομένων. Η παράμετρος main_table.name έχει τιμή Order_Details και στα αποτελέσματα βλέπουμε, την στήλη reference table, όπου είναι ο πίνακας συσχέτισης, και το main column, που σημαίνει ότι έχει ξένο κλειδί ως προς τον πίνακα συσχέτισης και την στήλη reference column που είναι το μοναδικό κλειδί του reference_table.



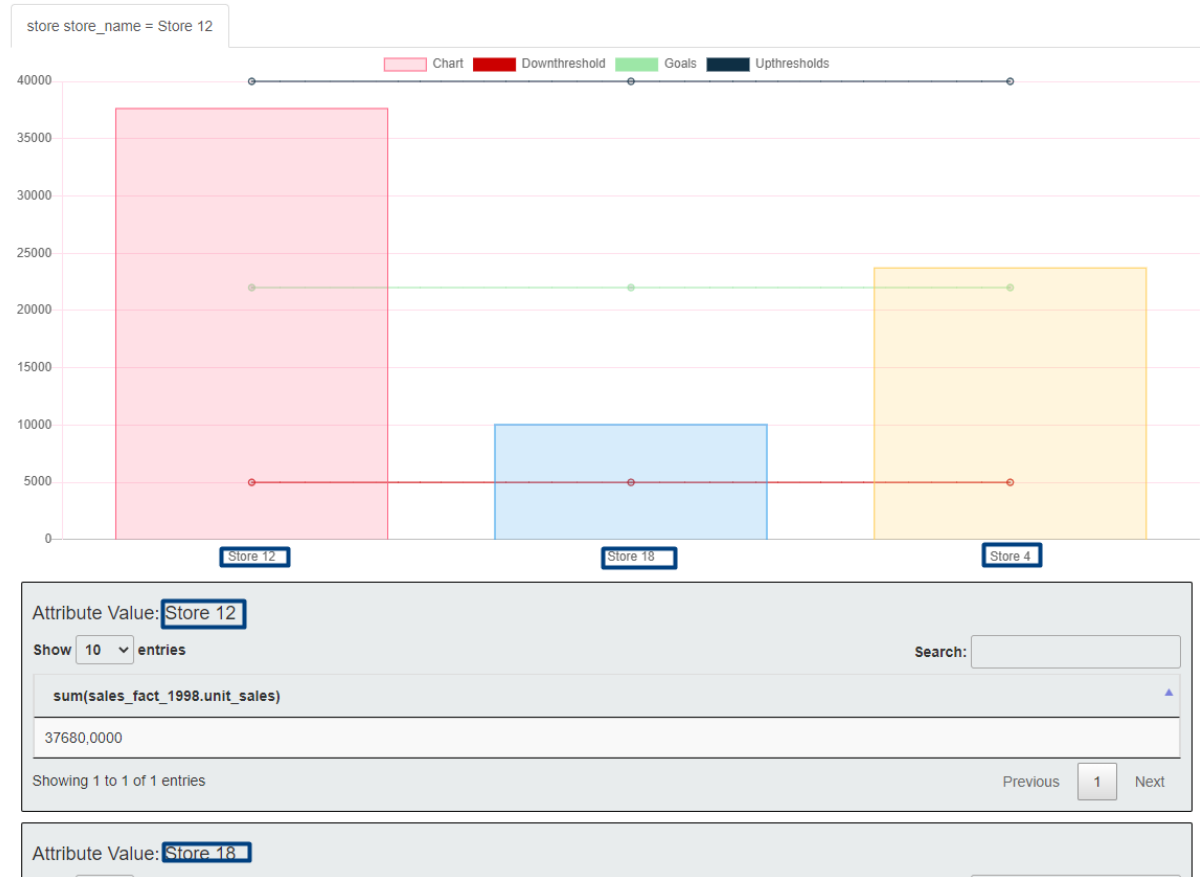
Σχήμα 3.7 Database από Fact Table εμφάνιση Constraint Tables

3.5.3 Εύρεση παρόμοιων τιμών για πίνακα διαστάσεων

Το παρακάτω παράδειγμα αφορά την περίπτωση που ο χρήστης έχει δηλώσει παράμετρο για κάποιο από τους πίνακες συσχέτισης. Κατά την αποθήκευση για αυτόν τον πίνακα έχει γίνει διερεύνηση σχετικά με το επίπεδο που πρόκειται να γίνει η ανάλυση. Στο παράδειγμά αυτό δηλώνεται ως βάση δεδομένων η foodmart, ο κύριος πίνακας sales_fact_1998 και ο χρήστης έχει δηλώσει στον πίνακα συσχέτισης store την παράμετρο store_name = Store 12 και σαν ανώτερο επίπεδο την store_state.

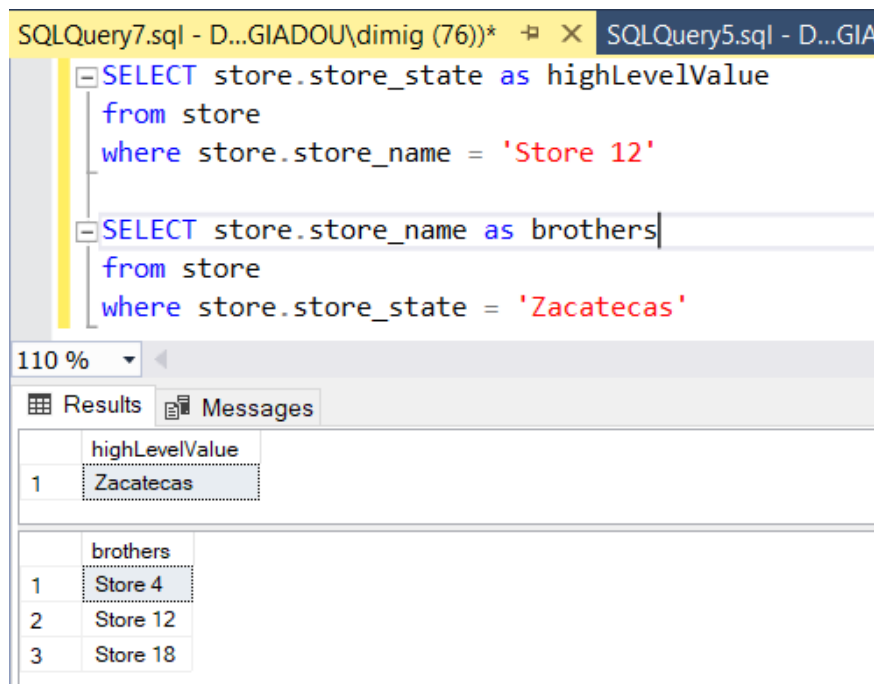
Στο σχήμα 3.8 βλέπουμε τα αποτελέσματα στη διεπαφή του χρήστη για την τιμή Store 12, για την οποία βρέθηκαν 3 αποτελέσματα αδέρφια Store13, Store 12, Store 18 με τα αποτελέσματα τους μέσω της εφαρμογής μιας συναθροιστικής συνάρτησης στο σχετικό μέτρο. Καθώς και το διάγραμμα για τα αποτελέσματα αδέρφια και τα φράγματα που έχει δηλώσει ο χρήστης στο KPI.

Results KPI



Σχήμα 3.8 UI από Αποτελέσματα για τιμές αδέρφια

Στο σχήμα 3.9 βλέπουμε την πρώτη ερώτηση, η οποία αφορά την εύρεση της τιμής από το ανώτερο επίπεδο με αποτέλεσμα, Zacatecas. Η δεύτερη ερώτηση αφορά την εύρεση αποτελεσμάτων αδέρφια, για τον πίνακα store, όταν το ανώτερο επίπεδο έχει τιμή Zacatecas, τα αποτελέσματα είναι Store 4, Store 12, Store 18.

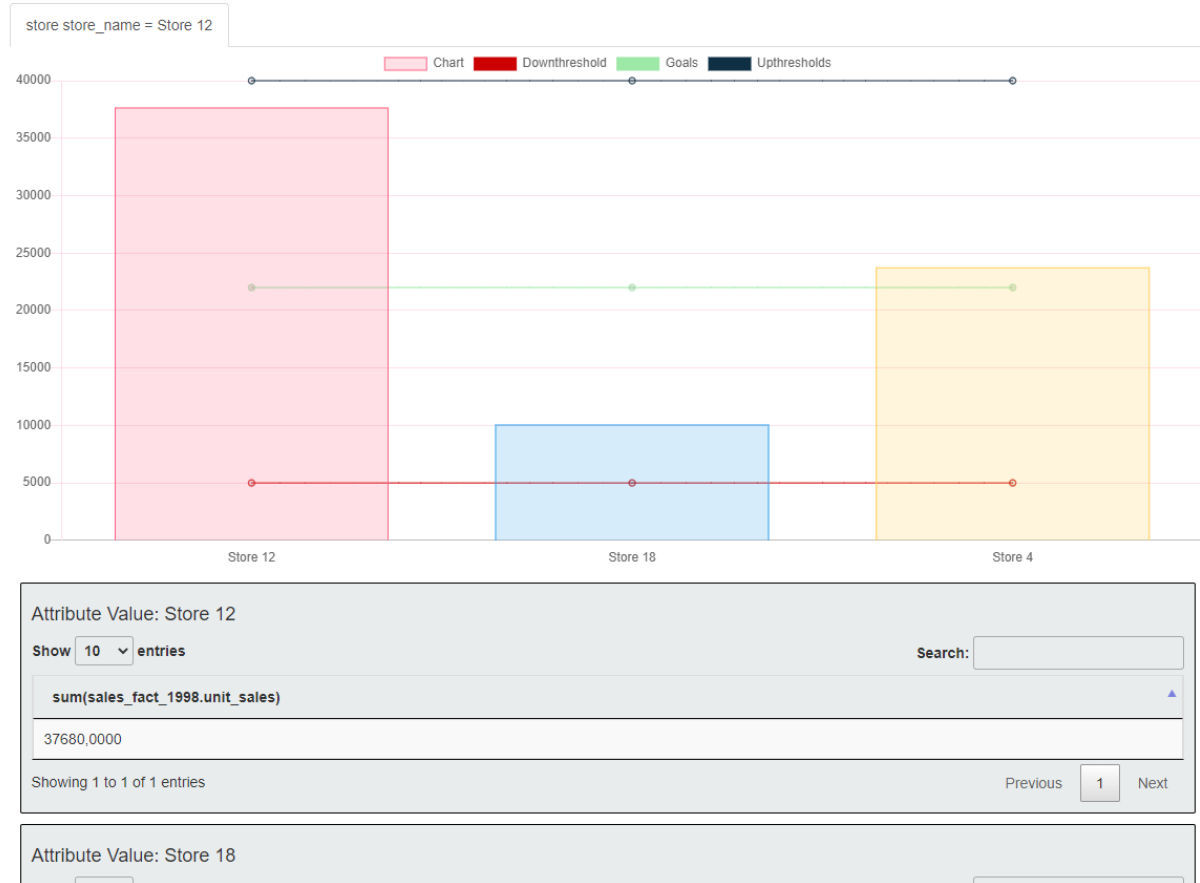


Σχήμα 3.9 Database Αποτελέσματα για τιμές αδέρφια

3.5.4 Αποτελέσματα με βάση συγκεκριμένο πίνακα διαστάσεων του Fact Table

Το παρακάτω παράδειγμα αφορά την περίπτωση που ο χρήστης για κάποιο από τους πίνακες συσχέτισης έχει δηλώσει παράμετρο. Κατά την αποθήκευση για τον πίνακα αυτό έχει γίνει διερεύνηση για ποιο επίπεδο θα γίνει η ανάλυση. Στο παράδειγμά αυτό δηλώνεται για τη βάση foodmart ο πίνακας sales_fact_1998 και ο χρήστης έχει δηλώσει στον πίνακα συσχέτισης store παράμετρο Store 12 και ανώτερο επίπεδο store_state.

Results KPI



Σχήμα 3.10 UI από Αποτελέσματα ανάλυσης

Ενώ η εκτέλεση της ερώτησης από τη βάση δεδομένων στον πίνακα store για την παράμετρο store_name με τιμή Store 4 απεικονίζεται στο σχήμα 3.11, η ερώτηση για την τιμή Store 18 απεικονίζεται στο σχήμα 3.12 και η ερώτηση για την τιμή Store 12 στο σχήμα 3.13

SQLQuery7.sql - D...\GIADOU\dimig (76))* X SQLQuery5.sql - D...\GIADOU\dimig

```

select sum(sales_fact_1998.unit_sales)
from sales_fact_1998
join store on sales_fact_1998.store_id = store.store_id
where store.store_name = 'Store 4'

```

110 %

Results Messages

	(No column name)
1	23752.0000

Σχήμα 3.11 Database Αποτελέσματα με τιμή Store 4

```

select sum(sales_fact_1998.unit_sales)
from sales_fact_1998
join store on sales_fact_1998.store_id = store.store_id
where store.store_name = 'Store 18'

```

110 %

Results Messages

	(No column name)
1	10070.0000

Σχήμα 3.12 Database Αποτελέσματα με τιμή Store 18

```

select sum(sales_fact_1998.unit_sales)
from sales_fact_1998
join store on sales_fact_1998.store_id = store.store_id
where store.store_name = 'Store 12'

```

110 %

Results Messages

	(No column name)
1	37680.0000

Σχήμα 3.13 Database Αποτελέσματα με τιμή Store 12

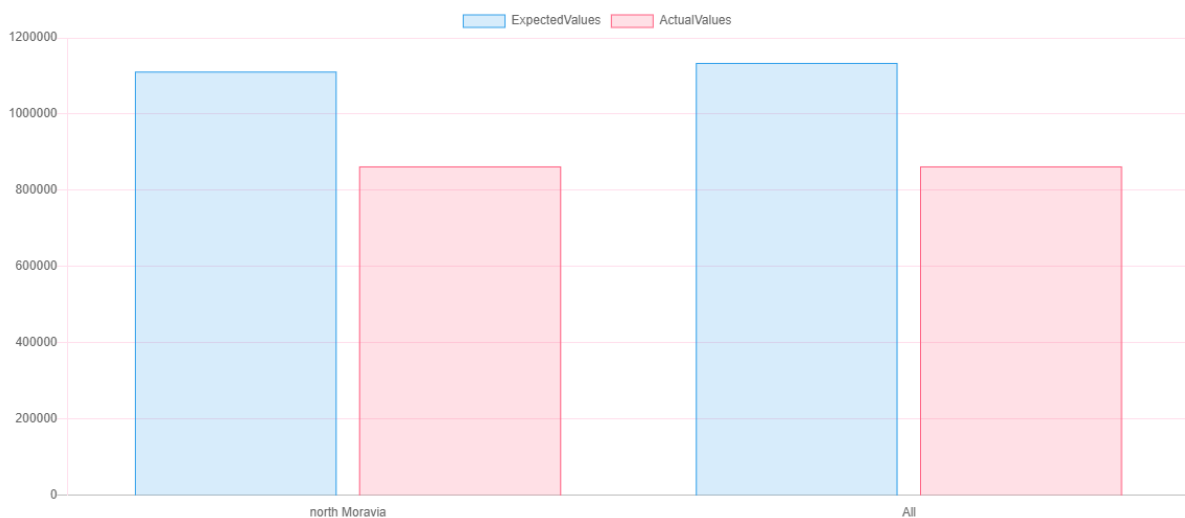
3.5.5 Σύγκριση της πραγματικής τιμής του KPI με την αναμενόμενη τιμή ως προς το ανώτερο επίπεδο και ως προς το σύνολο.

Το παρακάτω παράδειγμα αφορά τη σύγκριση της πραγματικής τιμής του KPI με την αναμενόμενη τιμή ως προς το ανώτερο επίπεδο και ως προς το σύνολο. Ο χρήστης στο συγκεκριμένο παράδειγμα έχει δηλώσει σαν βάση δεδομένων τη pkdd, σαν Fact Table τον πίνακα loan, έχει δηλώσει σαν φίλτρο και έτος ανάλυσης το 1994, από τον πίνακα account την παράμετρο district_name με τιμή Opava και ανώτερο επίπεδο region. Συνεπώς, στα μπλε διαγράμματα φαίνονται οι αναμενόμενες τιμές συγκριτικά με το προηγούμενο έτος και στα κόκκινα απεικονίζονται οι πραγματικές. Η πρώτη ομάδα αφορά τη σύγκριση ως προς το ανώτερο επίπεδο με τιμή north Moravia και η δεύτερη ως προς όλα. Παρατηρούμε πως η αναμενόμενη τιμή είναι μεγαλύτερη από την πραγματική, αυτό σημαίνει πως το έτος έρευνας έχει λιγότερες πωλήσεις σε σχέση με το προηγούμενο.

Expected Values

account district_name = Opava

Results for Year: 1994



Attribute Value: Opava		
Show 10 entries		
Search:		
Type of analysis	ActualResults	Expected
All	863044	1134474,87
north Moravia	863044	1111940,77

Σχήμα 3.14 Αποτέλεσμα στο UI για σύγκριση πραγματικών με αναμενόμενες τιμές

ΚΕΦΑΛΑΙΟ 4

ΥΛΟΠΟΙΗΣΗ

4.1 Πλατφόρμα και Προγραμματιστικά εργαλεία

4.2 Λεπτομέρειες Υλοποίησης

4.3 Αρχιτεκτονική Λογισμικού

Στην ενότητα αυτή θα περιγράψουμε λεπτομερώς την υλοποίηση και την τεχνολογία που χρησιμοποιήθηκε προκειμένου να κατασκευαστεί το σύστημα.

4.1 Πλατφόρμα και Προγραμματιστικά εργαλεία

Για την υλοποίηση του συστήματος χρησιμοποιήθηκε .Net Framework και η προγραμματιστική γλώσσα C#.

.NET Framework

Το .NET Framework [Micro2] αναπτύχθηκε από τη Microsoft και εκτελείται κυρίως στο λειτουργικό σύστημα των Windows. Περιλαμβάνει μια μεγάλη βιβλιοθήκη κλάσεων που ονομάζεται Framework Class Library (FCL) και παρέχει γλωσσική δια λειτουργικότητα σε διάφορες γλώσσες προγραμματισμού. Τα προγράμματα που γράφονται για το .NET Framework εκτελούνται σε ένα περιβάλλον λογισμικού που ονομάζεται Common Language Runtime (CLR). Το CLR είναι μια εικονική μηχανή εφαρμογών που παρέχει υπηρεσίες όπως η ασφάλεια, η διαχείριση μνήμης και ο χειρισμός εξαιρέσεων. Το FCL και το CLR αποτελούν μαζί το .NET Framework. Το FCL παρέχει τη διεπαφή χρήστη, την πρόσβαση σε δεδομένα, τη συνδεσιμότητα βάσεων δεδομένων, την κρυπτογραφία, την ανάπτυξη εφαρμογών ιστού, τους αριθμητικούς αλγορίθμους και τις επικοινωνίες δικτύου. Οι προγραμματιστές παράγουν λογισμικό συνδυάζοντας τον πηγαίο τους

κώδικα με το .NET Framework και άλλες βιβλιοθήκες. Η Microsoft παράγει επίσης ένα ολοκληρωμένο περιβάλλον ανάπτυξης λογισμικού .NET που ονομάζεται Visual Studio.

C#

Η C# [Micro3] μια γλώσσα προγραμματισμού. Δημιουργήθηκε από την Microsoft μέσα από την πλατφόρμα .NET Framework. Είναι μια από τις γλώσσες προγραμματισμού που δημιουργήθηκαν για την Κοινή Υποδομή Γλώσσας (Common Language Infrastructure). Ο κύριος σκοπός της γλώσσας είναι να είναι απλή αντικειμενοστραφής γλώσσα για γενική χρήση.

Microsoft SQL Server

Για βάση δεδομένων χρησιμοποιήθηκε Microsoft SQL Server [Micro4]. Ο SQL Server είναι μια σχεσιακή βάση δεδομένων, η οποία αναπτύσσεται από τη Microsoft. Οι κύριες γλώσσες που χρησιμοποιούνται είναι η T-SQL και η ANSI SQL. Η κύρια μονάδα αποθήκευσης στοιχείων είναι μια βάση δεδομένων, η οποία αποτελείται από μια συλλογή πινάκων και κώδικα.

Chart.js

Για τη απεικόνιση των γραφημάτων χρησιμοποιήθηκε το εργαλείο Chart.js [Chart1]. Το εργαλείο βοηθά την γραφική αναπαράσταση πληροφοριών και δεδομένων χρησιμοποιώντας γραφικά στοιχεία όπως είναι τα διαγράμματα, γραφήματα και ούτω καθεξής. Κατασκευάζονται τα διαγράμματα σε JavaScript και υπάρχουν πολλές δυνατότητες του στυλ και της λειτουργικότητας τους.

4.2 Λεπτομέρειες Υλοποίησης

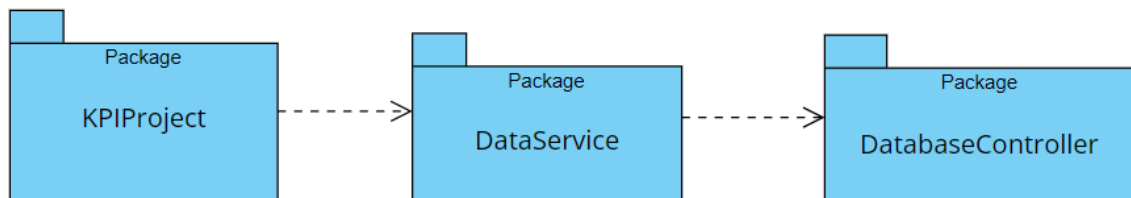
Το σύστημα αποτελείται από 3 εφαρμογές:

KPIProject: είναι ASP.NET [Micro5] Project και αφορά τη διεπαφή του χρήστη (UI). Το ASP.NET Core είναι προγραμματιστικό περιβάλλον της Microsoft που δημιουργήθηκε με σκοπό την υλοποίηση διαδικτυακών εφαρμογών.

DataService: είναι WCF (Windows Communication Foundation) [Micro6] αφορά τη διασύνδεση της διεπαφής του χρήστη με το DatabaseController (υλοποίηση βάσης). Το WCF είναι API στο .NET Framework για τη δημιουργία συνδέσμων μεταξύ εφαρμογών.

DatabaseController: είναι Class Library Application [Micro7] και περιέχει όλη την υλοποίηση της εφαρμογής. Σε αυτό γίνεται η διασύνδεση με τη βάση και δημιουργείται το Query για την εκτέλεση της ανάλυσης KPI.

4.3 Αρχιτεκτονική του Λογισμικού



Σχήμα 4.1 Package Diagram

DatabaseController

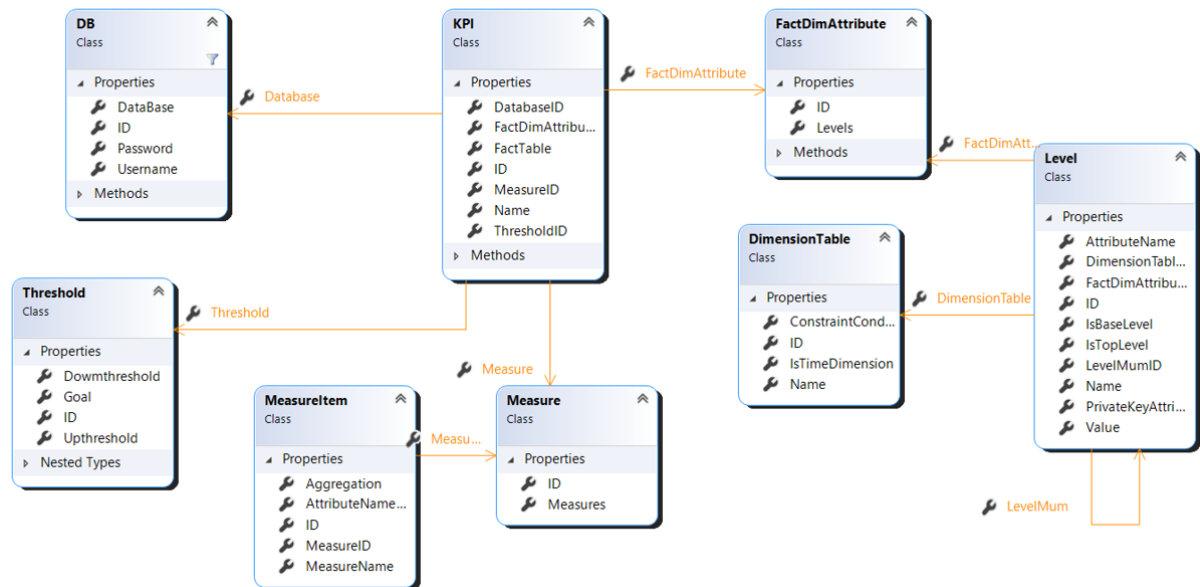
Το πακέτο DatabaseController περιέχει δύο πακέτα



Σχήμα 4.2 DatabaseController Package Diagram

Το πακέτο Controller περιέχει την κλάση QueryController, η οποία διαχειρίζεται όλες τις ερωτήσεις που θα εκτελεστούν προς τη Βάση Δεδομένων με βάση τα δεδομένα που έχουν οριστεί στο πακέτο Models.

Το πακέτο Models είναι δομημένο με την αρχιτεκτονική Object-Relational Mapping (ORM), η οποία, είναι μια τεχνική προγραμματισμού στην οποία χρησιμοποιείται ένας μηχανισμός περιγραφής των μεταδεδομένων για τη σύνδεση του αντικειμένου με μια σχεσιακή βάση δεδομένων. Συνεπώς, το πακέτο περιέχει όλους τους πίνακες που υπάρχουν στη βάση. Τα δεδομένα τους χρησιμοποιούνται για να εκτελέσουν οι ερωτήσεις προς τη βάση δεδομένων, ώστε με τα αποτελέσματα να γίνει η ανάλυση του KPI.



Σχήμα 4.3 Class Diagram of package Models

Το σχήμα των πινάκων είναι τέτοιο, ώστε να μπορεί να περιγράψει όλη τη μεταπληροφορία που αφορά ένα Star-Schema . Συγκεκριμένα:

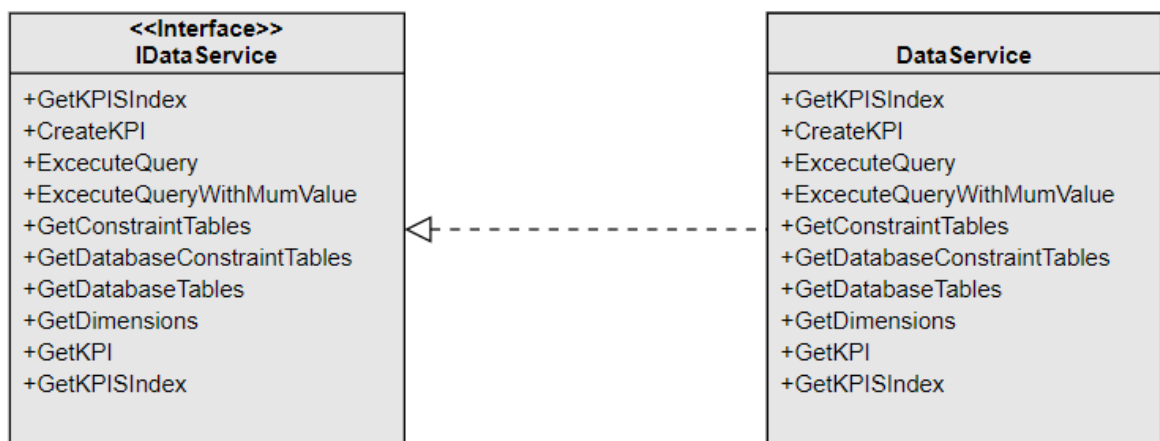
- **KPI:** πίνακας είναι ο βασικός περιέχει όλες τις συσχετίσεις με τους πίνακες DB, Measure, Threshold, FactDimAttribute καθώς και τον Fact Table που αφορά η ανάλυση.
- **DB:** περιέχει τα διαπιστευτήρια (credentials) για σύνδεση με την απομακρυσμένη βάση.
- **Measure:** περιέχει τις μετρήσεις για τα αποτελέσματα των ερωτήσεων στη βάση δεδομένων.

- **Threshold:** Περιέχει τα φράγματα (άνω/κάτω) του KPI καθώς και τιμή του στόχου
- **FactDimAttribute:** Αφορά τις συσχετίσεις του Fact Table
 - Ο πίνακας FactDimAttribute συσχετίζεται με τον πίνακα Level, ο οποίος διαθέτει τις πληροφορίες σε ποιο επίπεδο ιεραρχίας βρίσκεται, αν έχει κάποια παράμετρο που θα χρησιμοποιηθεί σαν φίλτρο κατά την εκτέλεση του KPI και αν περιέχεται άλλη κλάση που θα είναι η μητέρα του πίνακα συσχέτισης.
 - Ο πίνακας Level σχετίζεται με τον πίνακα DimensionTable, ο οποίος περιέχει τις πληροφορίες του πίνακα που συσχετίζεται.

Dataservice

Το πακέτο DataService περιέχει τις δυνατές κλήσεις προς τη βάση από την εφαρμογή της διεπαφής του χρήστη. Περιέχει τις κλάσεις:

- **IDataService.cs:** Αφορά τη διεπαφή (Interface) μεταξύ της υπηρεσίας και του πελάτη (client).
- **DataService.cs:** Αφορά την υλοποίηση της διεπαφής για τις κλήσεις που λαμβάνει.

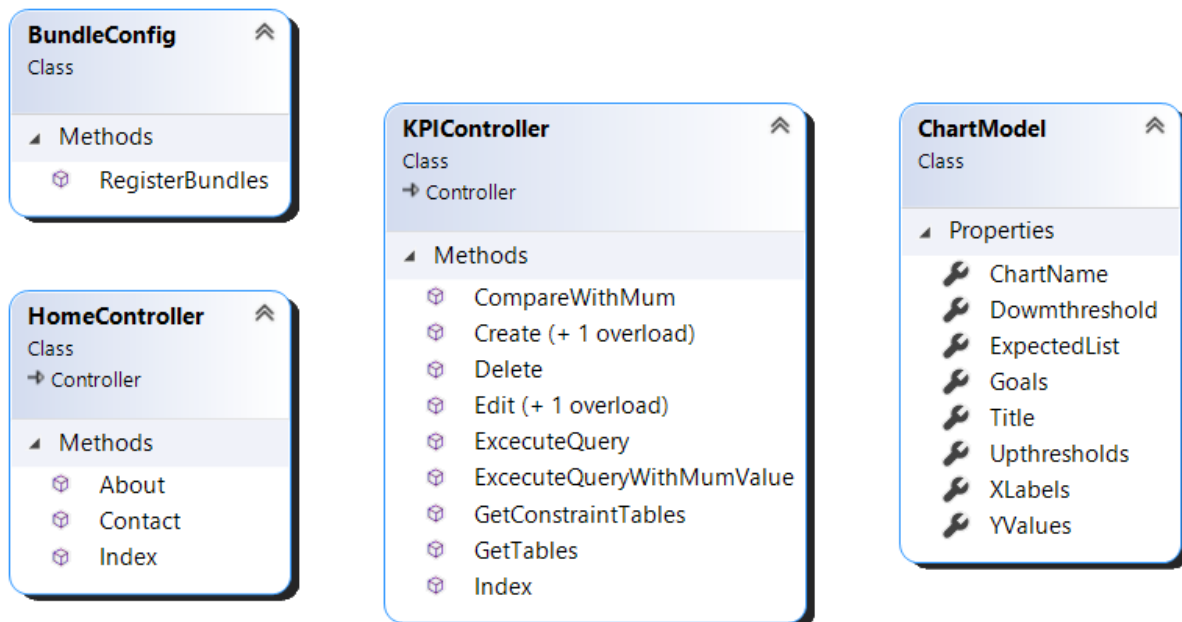


Σχήμα 4.4 Class Diagram πακέτου Data Service

KPIProject

Το πακέτο αυτό αφορά τη διεπαφή του χρήστη με την εφαρμογή. Τα υποπακέτα από τα οποία, αποτελείται είναι τα εξής:

- **App_Start:** από το πακέτο αυτό ξεκινάει η εφαρμογή και δημιουργεί τα Web Templates.
- **Content:** περιέχει όλα τα css αρχεία για το Style της εφαρμογής
- **Controllers:** το πακέτο αυτό επεξεργάζεται τα εισερχόμενα αιτήματα εκτελούν λειτουργίες και αποστέλλουν τα δεδομένα στο πακέτο Views ώστε να εμφανιστούν στον χρήστη.
- **Fonts:** περιέχει την γραμματοσειρά που χρησιμοποιεί η διεπαφή
- **Scripts:** περιέχει τα scripts που χρησιμοποιεί η διεπαφή του χρήστη
- **Views:** είναι η διεπαφή του χρήστη (UI) που εμφανίζει τα δεδομένα που προέρχονται από τους Controllers.



Σχήμα 4.5 Class Diagram όλου του πακέτου KPI Project

Κάποια από τα υποπακέτα δεν περιέχουν κλάσεις, όπως είναι το Content που περιέχει τα αρχεία css, το Fonts που περιέχει τα αρχεία για τη γραμματοσειρά, το πακέτο Script περιέχει τα Javascripts αρχεία και το υποπακέτο Views που περιέχει τα cshtml.

4.4 Έλεγχος αποτελεσμάτων

Για τον έλεγχο των αποτελεσμάτων έγινε ανάλυση της εκάστοτε βάσης δεδομένων, ώστε να δημιουργηθούν πειράματα. Αρχικά, για κάθε βήμα της υλοποίησης δημιουργήθηκαν χειρωνακτικά οι αναμενόμενες, “ορθές” ερωτήσεις στη βάση δεδομένων και εκτελέστηκαν. Στη συνέχεια έγινε έλεγχος αν οι ερωτήσεις που δημιουργήθηκαν από το σύστημα είναι οι αναμενόμενες, καθώς και αν τα αποτελέσματά τους είναι ίδια με αυτά των “ορθών” ερωτήσεων που εκτελέστηκαν στη βάση δεδομένων.

ΚΕΦΑΛΑΙΟ 5

ΠΕΙΡΑΜΑΤΑ

5.1 Σύγκριση KPI με παρόμοιες τιμές, εκτέλεση ερώτησης για κάθε παρόμοια τιμή

5.2 Σύγκριση KPI με παρόμοιες τιμές, εκτέλεση μια ερώτησης με τιμή το ανώτερο επίπεδο

5.3 Σύγκριση της πραγματικής τιμής του KPI με την αναμενόμενη τιμή ως προς το ανώτερο επίπεδο και ως προς το σύνολο

Στην ενότητα αυτή θα περιγράψουμε λεπτομερώς τα πειράματα που πραγματοποιήθηκαν για τις υλοποιήσεις που αφορούν τις συγκρίσεις των αποτελεσμάτων ενός KPI για τους χρόνους που απαιτούνται κατά την εκτέλεση τους.

Όλα τα πειράματα διεξήχθησαν σε Η/Υ με τα εξής χαρακτηριστικά:

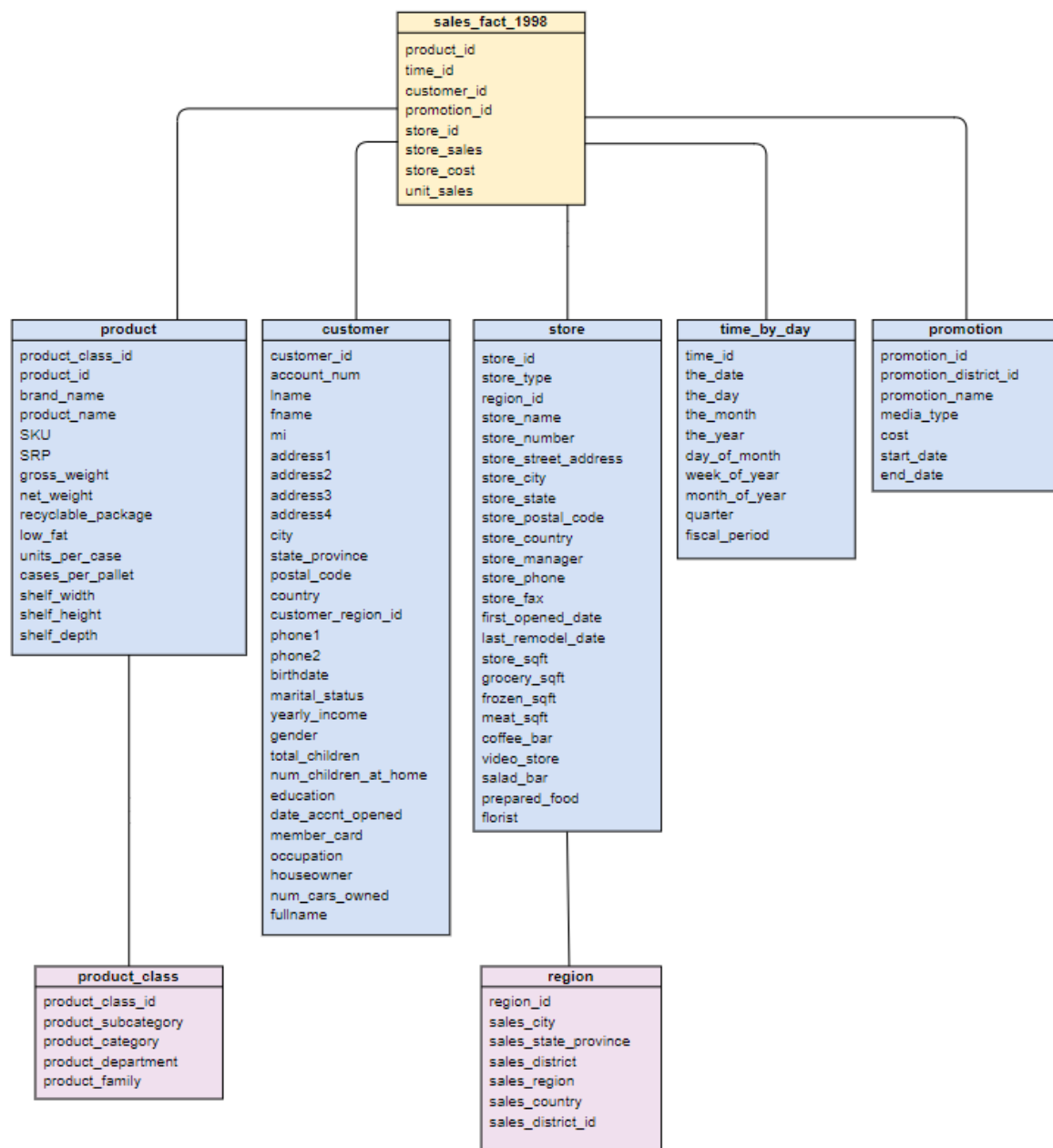
- CPU Intel(R) Core (TM) i7-10750H CPU @ 2.60GHz
- κύρια μνήμη 454 GB
- RAM 8 GB
- λειτουργικό σύστημα Windows 10.

5.1 Σύγκριση KPI με παρόμοιες τιμές, εκτέλεση ερώτησης για κάθε παρόμοια τιμή

5.1.1 Βάση Δεδομένων foodmart

Για την διεξαγωγή πειραμάτων χρησιμοποιήθηκε η βάση δεδομένων foodmart. Τα δεδομένα είναι αποθηκευμένα σε ένα σύνολο από πολυδιάστατους πίνακες, δηλαδή κύβους. Για τα πειράματα σαν Fact Table χρησιμοποιήθηκε ο πίνακας sales_fact_1998, ο οποίος διαθέτει 164558 εγγραφές και 5 πίνακες διατάσεων, Οι πίνακες διαστάσεων είναι

ο product, ο time, ο customer, ο promotion και ο store. Οι πίνακες product και store έχουν συσχέτιση με τους πίνακες product_class και region, αντίστοιχα.



Σχήμα 5.1 Πίνακες πειραμάτων στη βάση Foodmart

Οι τιμές των μετρήσεων είναι σε milliseconds και για το εκάστοτε πείραμα έχουν γίνει 5 εκτελέσεις.

Τα πειράματα που υλοποιήθηκαν έχουν κλιμακωτή αύξηση παραμέτρων ώστε να παρατηρήσουμε την συμπεριφορά του αλγορίθμου σε κάθε κατάσταση. Παρακάτω αναγράφονται τα πειράματα που διεξήχθησαν:

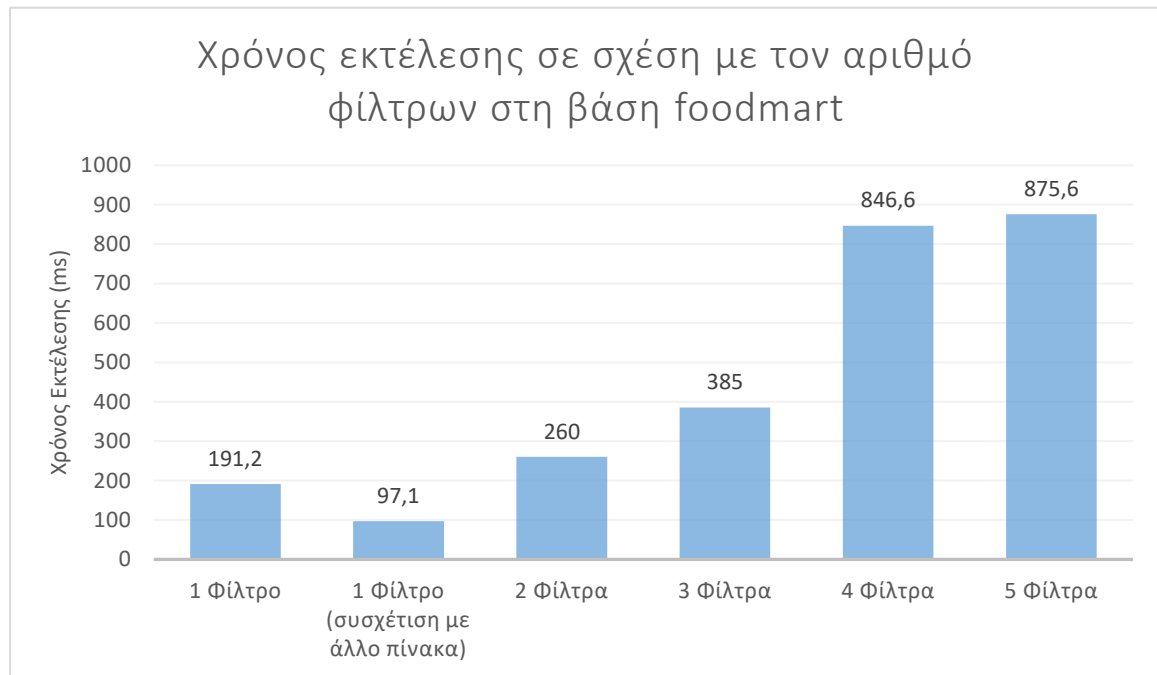
- **Πείραμα 1:** 1 φίλτρο -> 14 παρόμοιες τιμές
- **Πείραμα 2:** 2 φίλτρα -> 14, 9 παρόμοιες τιμές
- **Πείραμα 3:** 1 φίλτρο από join condition -> 8 παρόμοιες τιμές
- **Πείραμα 4:** 3 φίλτρα -> 14, 9, 3 παρόμοιες τιμές
- **Πείραμα 5:** 4 φίλτρα -> 14, 9, 3, 41 παρόμοιες τιμές
- **Πείραμα 6:** 5 φίλτρα -> 14, 9, 3, 41, 4 παρόμοιες τιμές

Για καθένα από τα φίλτρα που χρησιμοποιήθηκε αναγράφονται λεπτομερώς οι πληροφορίες παρακάτω:

- Για τον πίνακα store, η παράμετρος store_city με τιμή φίλτρου Los Angeles και ανώτερο επίπεδο την παράμετρο store_country. Αριθμός παιδιών του ανώτερου επιπέδου (αδέρφια της τιμής φίλτρου) = 14
- Για τον πίνακα product παράμετρος product_name με τιμή φίλτρου Great English Muffins και ανώτερο επίπεδο brand_name Αριθμός παιδιών του ανώτερου επιπέδου (αδέρφια της τιμής φίλτρου) = 9
- Για τον πίνακα store, δηλώθηκε 1 παράμετρος από τον χρήστη η region_id με τιμή φίλτρου 25, η οποία είναι ξένο κλειδί στον πίνακα region και σαν ανώτερο επίπεδο το region_state. Αριθμός παιδιών του ανώτερου επιπέδου (αδέρφια της τιμής φίλτρου) = 8
- Για τον πίνακα customer η παράμετρος state_province με τιμή φίλτρου CA και ανώτερο επίπεδο country. Αριθμός παιδιών του ανώτερου επιπέδου (αδέρφια της τιμής φίλτρου) = 3
- Για πίνακα promotion η παράμετρος promotion_name με τιμή φίλτρου Save-It Sale και ανώτερο επίπεδο media_type. Αριθμός παιδιών του ανώτερου επιπέδου (αδέρφια της τιμής φίλτρου) = 41
- Για πίνακα time_by_day η παράμετρος quarter με τιμή φίλτρου Q1 και ανώτερο επίπεδο the_year. Αριθμός παιδιών του ανώτερου επιπέδου (αδέρφια της τιμής φίλτρου) = 4

Στο παρακάτω γράφημα (σχήμα 5.2) γίνεται αναπαράσταση των χρόνων (κατά μέσο όρο 5 εκτελέσεων) για πειράματα από 1 μέχρι 5 φίλτρα. Παρατηρούμε πως όσο

αυξάνονται τα φίλτρα τόσο αυξάνεται και ο χρόνος εκτέλεσης για την διεξαγωγή των αυτόματων αναφορών ενός KPI με παρόμοιες τιμές. Ο λόγος που προκαλεί την αύξηση στον χρόνο εκτέλεσης, είναι οι επιπλέον ερωτήσεις που υποβάλλονται, μία για κάθε φίλτρο.



Σχήμα 5.2 Σχεδιάγραμμα για χρόνο εκτέλεσης σε σχέση με τον αριθμό φίλτρων στη βάση foodmart

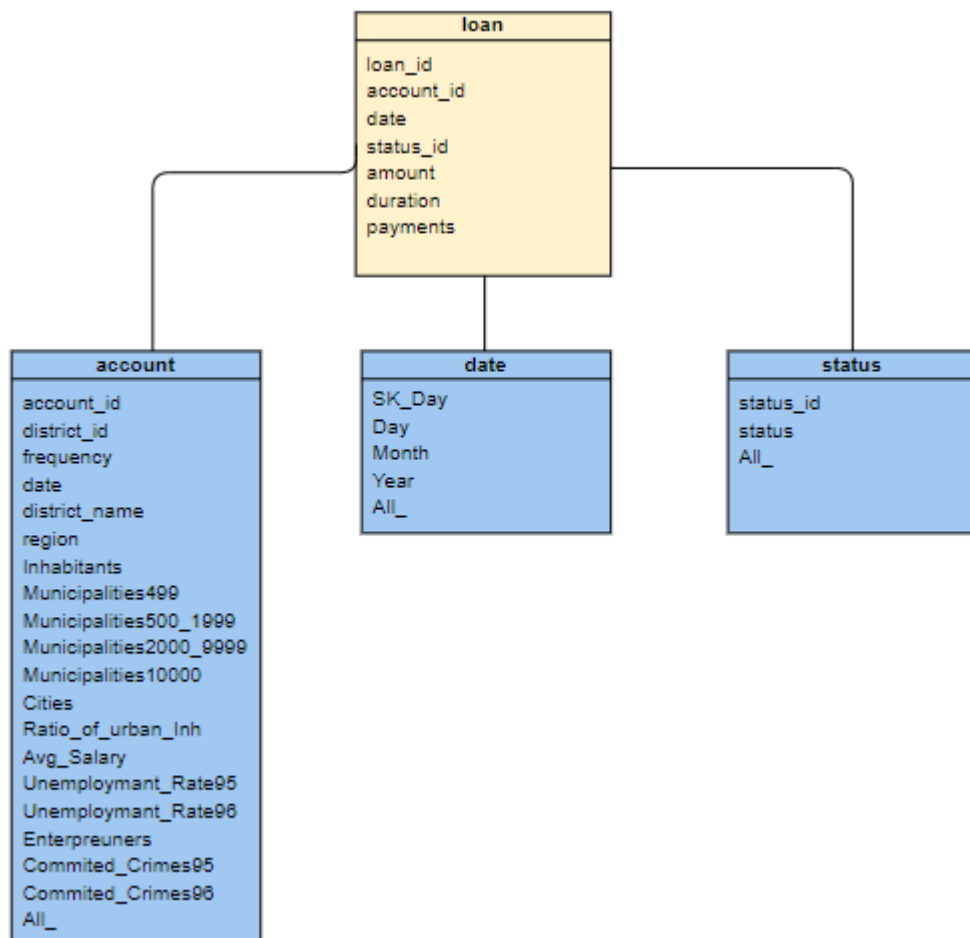
Στο παρακάτω γράφημα (σχήμα 5.3) γίνεται απεικόνιση των χρόνων που χρειάστηκε για να γίνει η εκτέλεση των παρόμοιων τιμών για κάθε φίλτρο, των πινάκων ανάλογα με τον αριθμό παιδιών που διαθέτει το καθένα. Παρατηρούμε, πως για μεγαλύτερο αριθμό παρόμοιων τιμών (τιμές αδέρφια) είναι μεγαλύτερος και ο χρόνος εκτέλεσης. Ο λόγος που προκαλεί την αύξηση στον χρόνο εκτέλεσης, είναι οι το πλήθος των ερωτήσεων που υποβάλλονται, μία για κάθε φίλτρο.



Σχήμα 5.3 Σχεδιάγραμμα για χρόνους εκτέλεσης σε σχέση με το πλήθος παρόμοιων τιμών (βάση foodmart)

5.1.2 Βάση Δεδομένων rkd99

Για να μετρήσουμε την επίπτωση του όγκου των δεδομένων στο χρόνο εκτέλεσης, κατασκευάσαμε τεχνητά δεδομένα για το ίδιο σχήμα. Για τη διεξαγωγή πειραμάτων χρησιμοποιήθηκε η βάση δεδομένων rkd99. Τα δεδομένα είναι αποθηκευμένα σε ένα σύνολο από πολυδιάστατους πίνακες, δηλαδή κύβους. Για τα πειράματα σαν Fact Table χρησιμοποιήθηκε ο πίνακας loan και υπάρχουν 3 πίνακες διατάσεων. Οι πίνακες διαστάσεων είναι ο account, ο date και ο status. Για την ανάλυση χρησιμοποιήθηκε ο πίνακας loan με 100.00 και 1.000.000 εγγραφές.



Σχήμα 5.4 Πίνακες πειραμάτων στη βάση pkdd99

Οι χρόνοι των μετρήσεων είναι σε milliseconds και για το εκάστοτε πείραμα έχουν γίνει 5 εκτελέσεις.

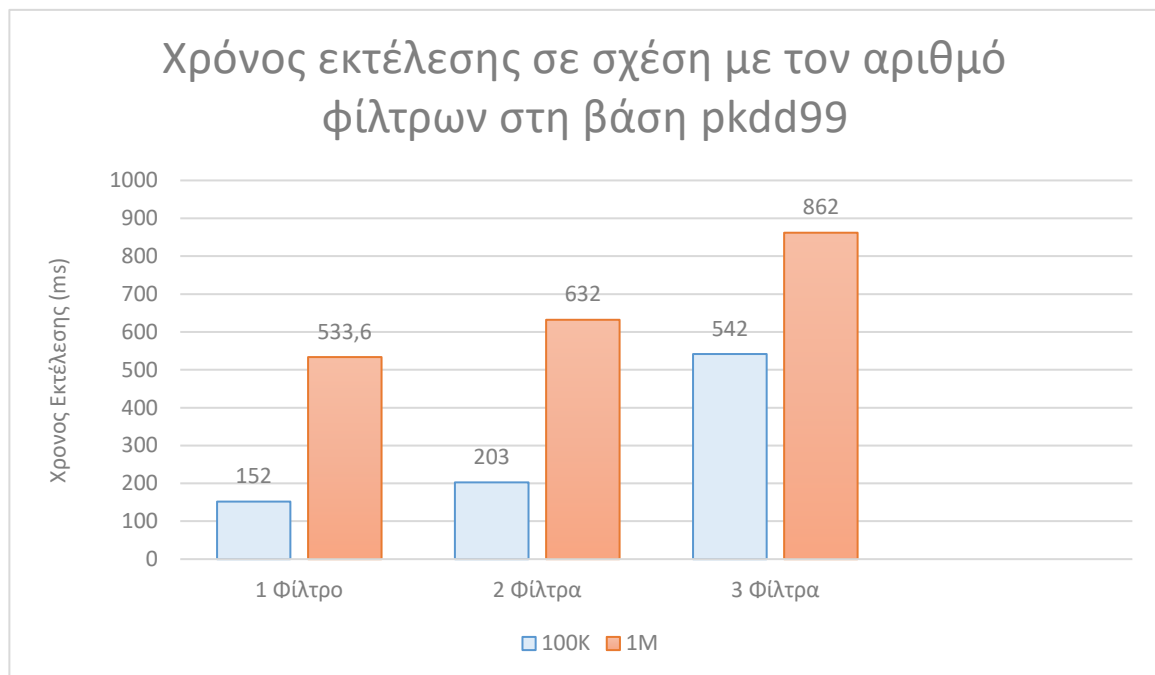
Τα πειράματα που υλοποιήθηκαν έχουν κλιμακωτή αύξηση παραμέτρων ώστε να παρατηρήσουμε την συμπεριφορά του αλγορίθμου σε κάθε κατάσταση. Παρακάτω αναγράφονται τα πειράματα που διεξήχθησαν:

- **Πείραμα 1:** 1 φίλτρο -> 11 παρόμοιες τιμές
- **Πείραμα 2:** 2 φίλτρα -> 11, 12 παρόμοιες τιμές
- **Πείραμα 3:** 3 φίλτρα -> 11, 12, 5 παρόμοιες τιμές

Για καθένα από τα φίλτρα που χρησιμοποιήθηκε αναγράφονται λεπτομερώς οι πληροφορίες παρακάτω:

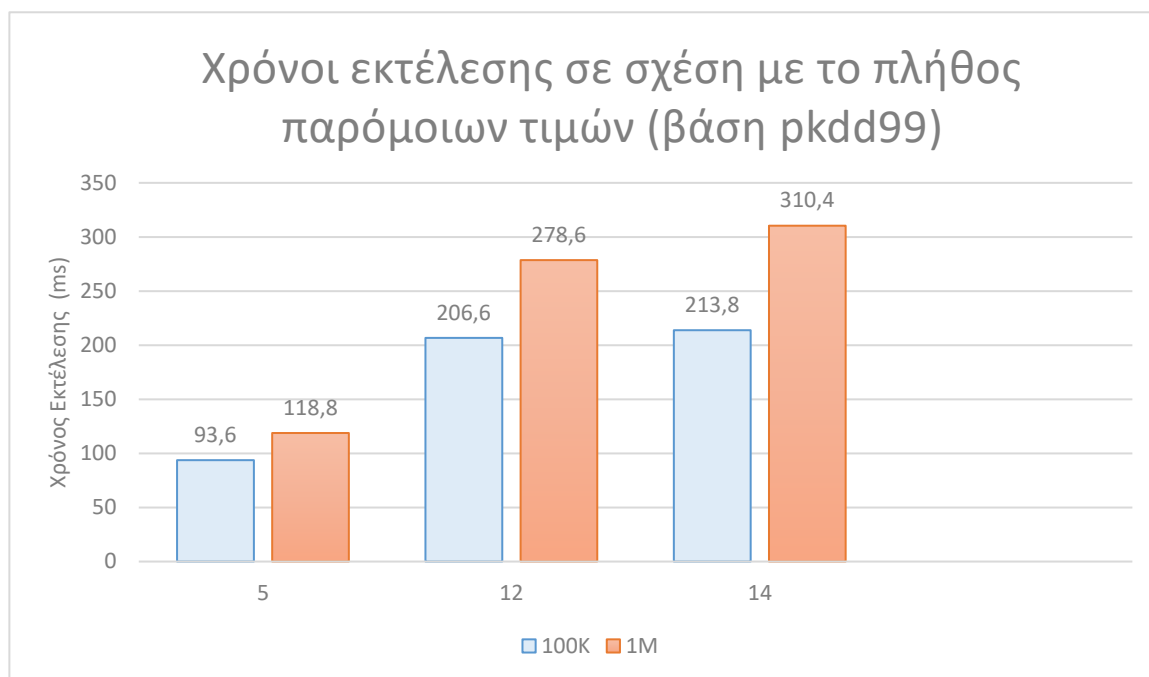
- για τον πίνακα διαστάσεων account, η παράμετρος district_name με τιμή Orana και ανώτερο επίπεδο την παράμετρο region. Αριθμός παιδιών του ανώτερου επιπέδου (αδέρφια της τιμής φίλτρου) = 11
- Για τον πίνακα date παράμετρος Month με τιμή 1994-01 και ανώτερο επίπεδο Year. Αριθμός παιδιών του ανώτερου επιπέδου (αδέρφια της τιμής φίλτρου) = 12
- Για τον πίνακα status παράμετρος status με τιμή Other status και ανώτερο επίπεδο All_. Αριθμός παιδιών του ανώτερου επιπέδου (αδέρφια της τιμής φίλτρου) = 5

Στο παρακάτω γράφημα (σχήμα 5.5) γίνεται αναπαράσταση των χρόνων κατά μέσο όρο 5 εκτελέσεων για πειράματα από 1 μέχρι 3 φίλτρα. Το γαλάζιο γράφημα αφορά τις μετρήσεις με 100.000 εγγραφές ενώ το πορτοκαλί αφορά 1.000.000. Παρατηρούμε πως όσο αυξάνονται τα φίλτρα τόσο αυξάνεται και ο χρόνος εκτέλεσης για την διεξαγωγή των αυτόματων αναφορών ενός KPI με παρόμοιες τιμές. Επιπλέον, παρατηρούμε πως όσο μεγαλύτερος είναι ο αριθμός των εγγραφών στον Fact Table τόσο περισσότερο αυξάνεται και ο χρόνος εκτέλεσης που απαιτείται.



Σχήμα 5.5 Σχεδιάγραμμα για χρόνο εκτέλεσης σε σχέση με τον αριθμό φίλτρων στη βάση pkdd99

Στο παρακάτω γράφημα (σχήμα 5.6) γίνεται απεικόνιση των χρόνων που χρειάστηκε για να γίνει η εκτέλεση των παρόμοιων τιμών για κάθε φίλτρο, των πινάκων ανάλογα με τον αριθμό παιδιών που διαθέτει το καθένα. Το γαλάζιο γράφημα αφορά τις μετρήσεις με 100.000 εγγραφές ενώ το πορτοκαλί αφορά 1.000.000. Παρατηρούμε, πως για μεγαλύτερο αριθμό παρόμοιων τιμών (τιμές αδέρφια) είναι μεγαλύτερος και ο χρόνος εκτέλεσης. Επιπλέον, παρατηρούμε πως όσο μεγαλύτερος είναι ο αριθμός των εγγραφών στον Fact Table τόσο περισσότερο αυξάνεται και ο χρόνος εκτέλεσης που απαιτείται για το εκάστοτε φίλτρο.



Σχήμα 5.6 Σχεδιάγραμμα για χρόνους εκτέλεσης σε σχέση με το πλήθος παρόμοιων τιμών (βάση rkdd99)

5.2 Σύγκριση KPI με παρόμοιες τιμές, εκτέλεση μια ερώτησης με τιμή το ανώτερο επίπεδο

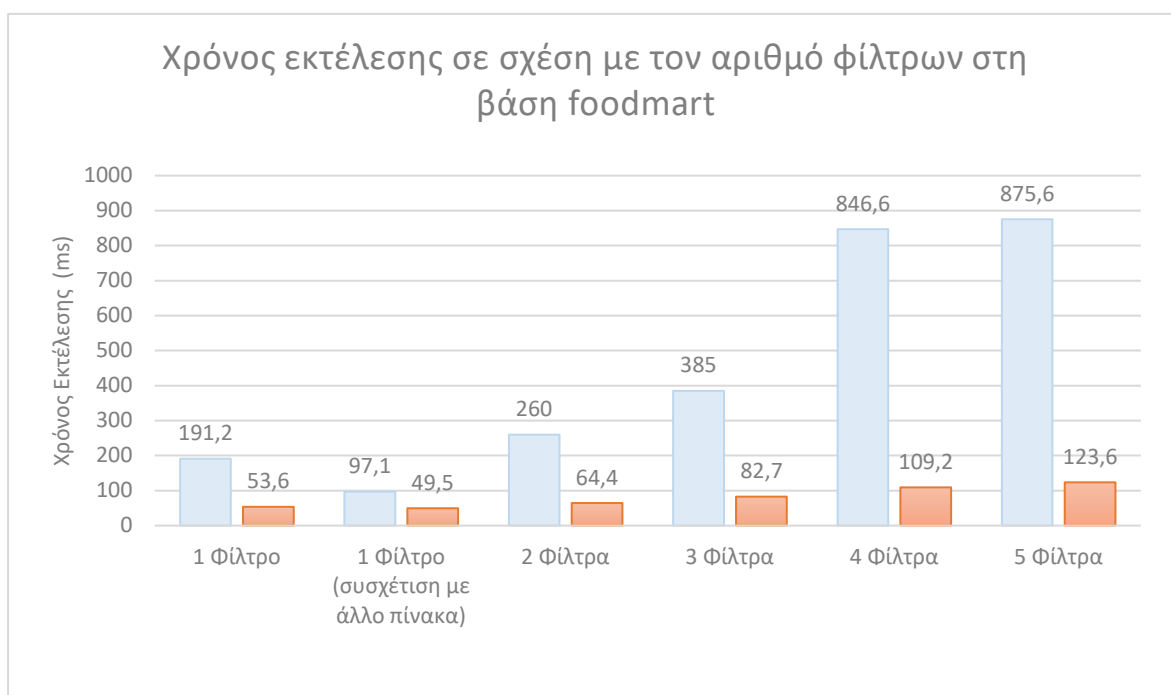
Για την εκτέλεση ενός KPI για τα αποτελέσματα των παρόμοιων τιμών με τα ποσά των μετρήσεων έχουν δημιουργηθεί δύο εναλλακτικές υλοποιήσεις. Η πρώτη υλοποίηση

αφορά την εκτέλεση μόνο μιας ερώτησης στη βάση δεδομένων, ώστε να ληφθούν τα αποτελέσματα των παρόμοιων τιμών και των μετρήσεων. Η δεύτερη υλοποίηση πραγματοποιεί μια σειρά από ερωτήσεις στη βάση δεδομένων, προκειμένου να ληφθούν τα αποτελέσματα των παρόμοιων τιμών με τις μετρήσεις. Τα παρακάτω πειράματα αφορούν την σύγκριση των χρόνων που απαιτείται για τις δύο εναλλακτικές υλοποιήσεις που έχουν πραγματοποιηθεί.

Οι τιμές των μετρήσεων είναι σε milliseconds και για το εκάστοτε πείραμα έχουν γίνει 5 εκτελέσεις

5.2.1 Βάση Δεδομένων foodmart

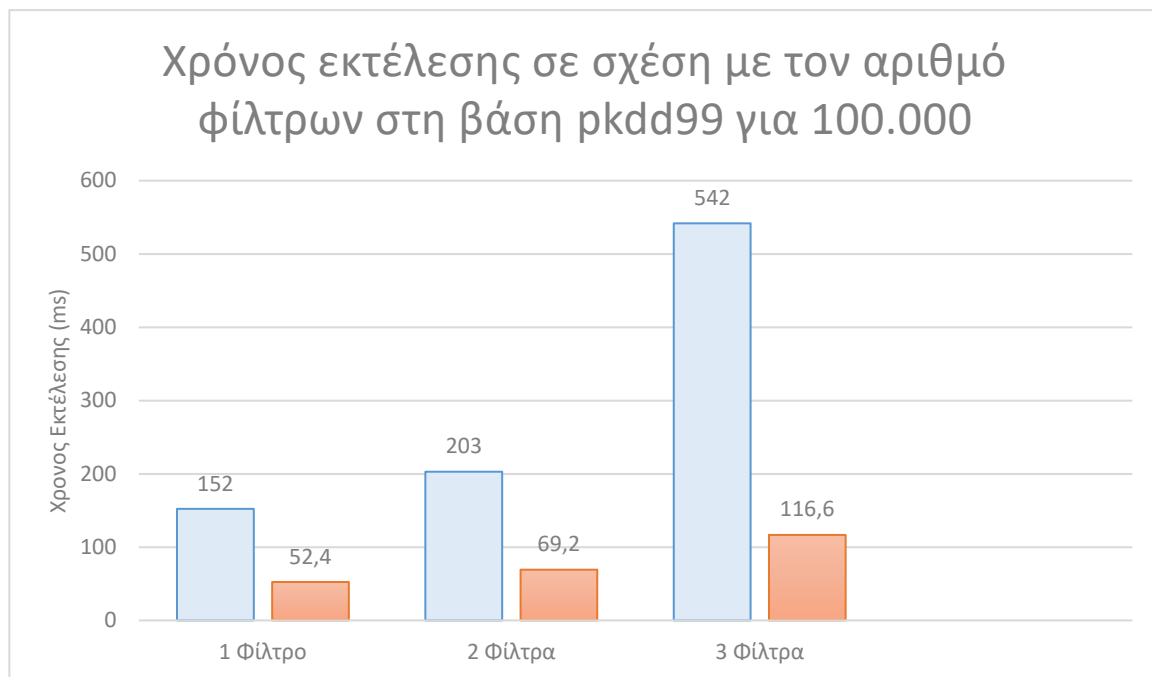
Στο παρακάτω γράφημα (σχήμα 5.7) γίνεται αναπαράσταση των χρόνων (κατά μέσο όρο 5 εκτελέσεων) για πειράματα από 1 μέχρι 5 φίλτρα. Με γαλάζιο χρώμα είναι τα γραφήματα για την ξεχωριστή εκτέλεση των παρόμοιων τιμών, ενώ με πορτοκαλί είναι τα διαγράμματα για την υλοποίηση που δημιουργεί μια ερώτηση για τα αποτελέσματα των παρόμοιων τιμών με τα ποσά της μέτρησης. Παρατηρούμε πως όσο αυξάνονται τα φίλτρα τόσο αυξάνεται και ο χρόνος εκτέλεσης για την διεξαγωγή των αυτόματων αναφορών ενός KPI με παρόμοιες τιμές. Ωστόσο, η διαφορά στους χρόνους μεταξύ των δυο υλοποιήσεων είναι πολύ μεγάλη: αυτό οφείλεται στον χρόνο που απαιτείται για την κάθε ξεχωριστή εκτέλεση των παρόμοιων τιμών για τα αποτελέσματα της μέτρησης σε αντίθεση με την εκτέλεση μιας ερώτησης που τα αποτελέσματα θα υπολογιστούν από τη μια εκτέλεση.



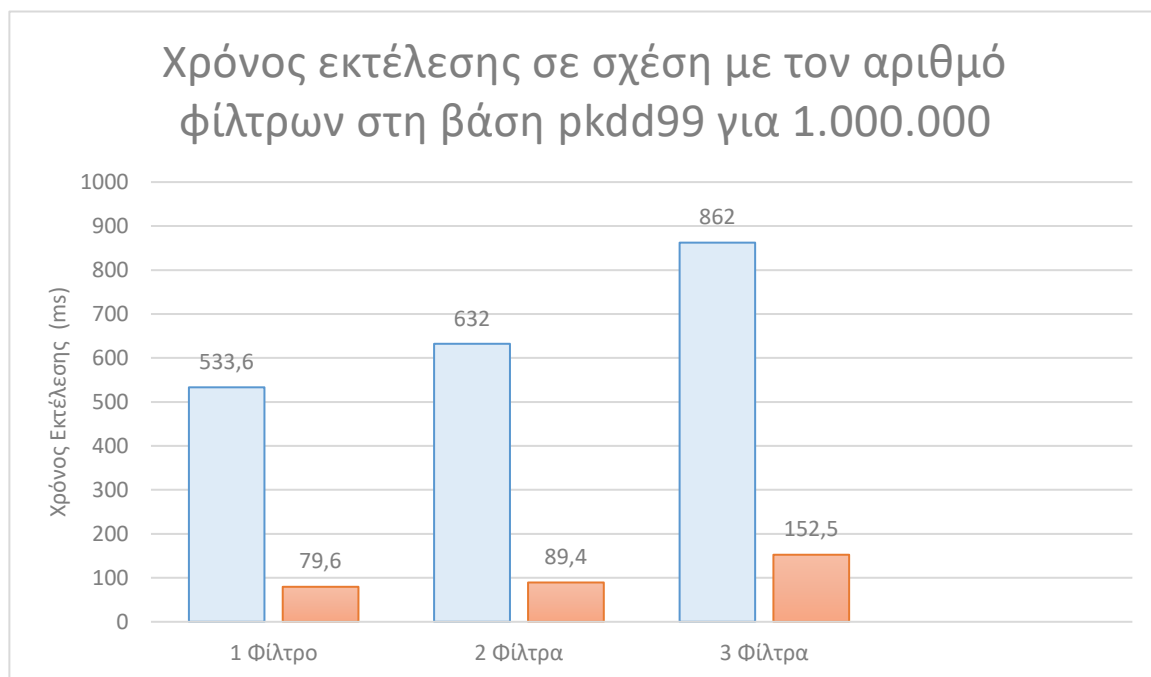
Σχήμα 5.7 Σχεδιάγραμμα για χρόνο εκτέλεσης σε σχέση με τον αριθμό φίλτρων στη βάση foodmart

5.2.2 Βάση Δεδομένων rkd99

Στο παρακάτω γραφήματα γίνεται αναπαράσταση των χρόνων κατά μέσο όρο 5 εκτελέσεων για πειράματα από 1 μέχρι 3 φίλτρα. Το πρώτο (σχήμα 5.8) αφορά μετρήσεις για 100.000 εγγραφές ενώ το δεύτερο (σχήμα 5.9) για 1.000.000. Το γαλάζιο γράφημα αφορά τις την υλοποίηση για ξεχωριστές εκτελέσεις ενώ το πορτοκαλί αφορά την εκτέλεση μιας ερώτησης. Παρατηρούμε πως όσο αυξάνονται τα φίλτρα τόσο αυξάνεται και ο χρόνος εκτέλεσης για την διεξαγωγή των αυτόματων αναφορών ενός KPI με παρόμοιες τιμές. Ωστόσο, όπως στην περίπτωση του σχήματος 5.1.5 τα αποτελέσματα είναι τα αναμενόμενα, η διαφορά στους χρόνους μεταξύ των δυο υλοποιήσεων είναι πολύ μεγάλη.



Σχήμα 5.8 Σχεδιάγραμμα για χρόνο εκτέλεσης σε σχέση με τον αριθμό φίλτρων στη βάση για 100.000 εγγραφές



Σχήμα 5.9 Σχεδιάγραμμα για χρόνο εκτέλεσης σε σχέση με τον αριθμό φίλτρων στη βάση για 1.000.000 εγγραφές

5.3 Σύγκριση της πραγματικής τιμής του KPI με την αναμενόμενη τιμή ως προς το ανώτερο επίπεδο και ως προς το σύνολο

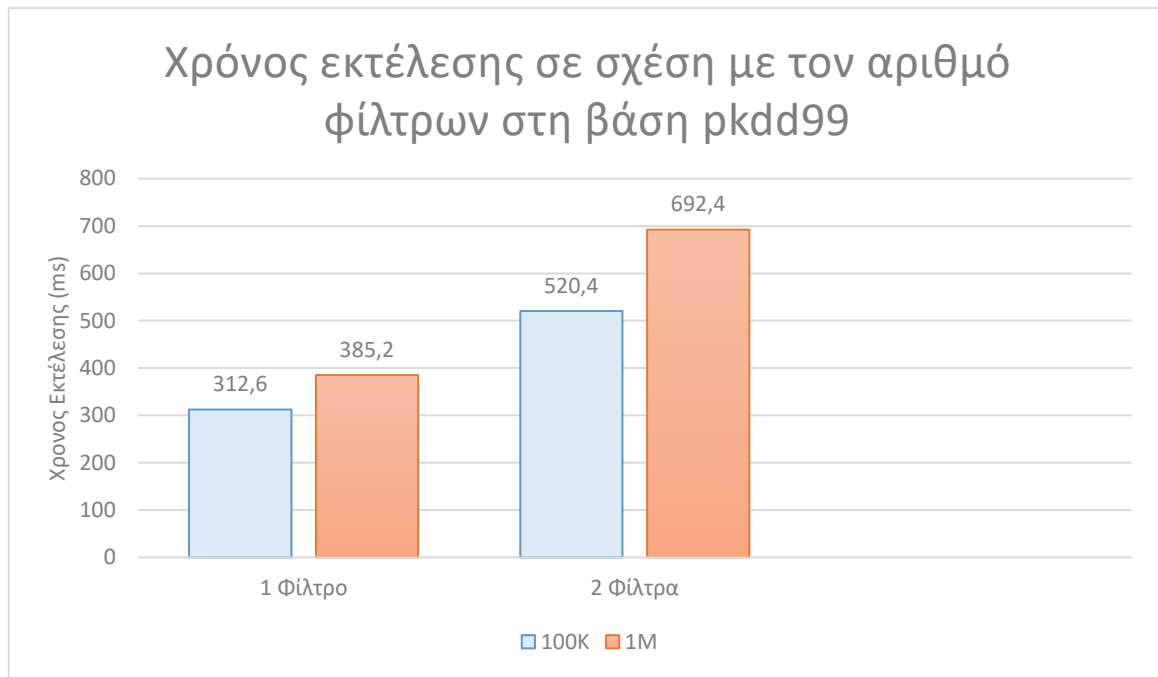
Στο συγκεκριμένο πείραμα αποτιμούμε τους αλγορίθμους 13 για το ανώτερο επίπεδο και αλγόριθμο 18 το γενικό σύνολο για τη σύγκριση ως προς τον χρόνο σε σχέση με το μέγεθος της βάσης δεδομένων και τον αριθμό των φίλτρων. Για τη διεξαγωγή πειραμάτων χρησιμοποιήθηκε η βάση δεδομένων pkd99. Τα δεδομένα είναι αποθηκευμένα σε ένα σύνολο από πολυδιάστατους πίνακες, δηλαδή κύβους. Για τα πειράματα, σαν Fact Table χρησιμοποιήθηκε ο πίνακας loan και υπάρχουν 3 πίνακες διαστάσεων. Οι πίνακες διαστάσεων είναι ο account, ο date και ο status. Για την ανάλυση χρησιμοποιήθηκε ο πίνακας loan με 100.000 και 1.000.000 εγγραφές.

Οι χρόνοι των μετρήσεων είναι σε milliseconds και για το εκάστοτε πείραμα έχουν γίνει 5 εκτελέσεις.

Για καθένα από τα φίλτρα που χρησιμοποιήθηκαν αναγράφονται λεπτομερώς οι πληροφορίες παρακάτω:

- Για τον πίνακα διαστάσεων account, η παράμετρος district_name με τιμή Orava και ανώτερο επίπεδο την παράμετρο region.
- Για τον πίνακα date παράμετρος Year με τιμή 1994.
- Για τον πίνακα status παράμετρος status με τιμή Running Contract/OK και ανώτερο επίπεδο All_.

Στο παρακάτω γράφημα γίνεται αναπαράσταση των χρόνων κατά μέσο όρο 5 εκτελέσεων για πειράματα. Το γαλάζιο γράφημα αφορά τις μετρήσεις με 100.000 εγγραφές, ενώ το πορτοκαλί αφορά 1.000.000. Παρατηρούμε πως όσο αυξάνονται τα φίλτρα τόσο αυξάνεται και ο χρόνος εκτέλεσης για την διεξαγωγή των αυτόματων αναφορών ενός KPI για συγκρίσεις των πραγματικών με τις αναμενόμενες τιμές. Επιπλέον, παρατηρούμε πως όσο μεγαλύτερος είναι ο αριθμός των εγγραφών στον Fact Table τόσο περισσότερο αυξάνεται και ο χρόνος εκτέλεσης που απαιτείται.



Σχήμα 5.10 Σχεδιάγραμμα για χρόνο εκτέλεσης σε σχέση με τον αριθμό φίλτρων στη βάση pkdd99 για τις αναμενόμενες τιμές

ΚΕΦΑΛΑΙΟ 6

ΣΥΜΠΕΡΑΣΜΑΤΑ

6.1 Σύνοψη αποτελεσμάτων

6.2 Μελλοντικές Επεκτάσεις

Στην ενότητα αυτή αναφέρονται η σύνοψη των αποτελεσμάτων καθώς και οι μελλοντικές επεκτάσεις.

6.1 Σύνοψη αποτελεσμάτων

Για τη μεταπτυχιακή εργασία ειδίκευσης υλοποιήθηκε ένα ολοκληρωμένο σύστημα το οποίο αφορά τη δημιουργία και αποτίμηση Κρίσιμων Δεικτών Απόδοσης που χρησιμοποιούνται για την εκτίμηση της αποτελεσματικότητας ενός οργανισμού και την παρακολούθηση των επιδόσεων τους. Διαθέτει μια βάση δεδομένων και τη διεπαφή του χρήστη, ώστε με τη δημιουργία Κρίσιμων Δεικτών Απόδοσης (KPIs), το οποίο αποτιμά μια τιμή-μέτρησης σε σχέση με ένα υποσύνολο των δεδομένων, για το οποίο, ο αναλυτής έχει δώσει και ένα μέτρο της αναμενόμενης τιμής του, και ένα τρόπο χαρακτηρισμού του αποτελέσματος σε σχέση με την αναμενόμενη τιμή. Κατά τη δημιουργία ενός KPI ο χρήστης καταχωρεί τα στοιχεία της βάσης δεδομένων, τον κύριο πίνακα ανάλυσης (Fact Table), τα φίλτρα στους πίνακες διαστάσεων με τους οποίους συσχετίζεται, τη μέτρηση και τα φράγματα. Για την παρακολούθηση της απόδοσης του οργανισμού ο χρήστης μπορεί να την ελέγξει από δύο είδη αναφορών που προκύπτουν από μια σειρά ερωτήσεων προς τη βάση δεδομένων. Τα είδη των αναφορών είναι τα εξής:

Στην πρώτη αναφορά για κάθε φίλτρο που έχει δηλώσει ο χρήστης στους πίνακες διαστάσεων θα βρεθούν και οι παρόμοιες τιμές και θα υπολογιστούν τα αποτελέσματα του συναθροιστικού ποσού της μέτρησης, ώστε να γίνουν οι μεταξύ τους συγκρίσεις. Τα αποτελέσματα απεικονίζονται σε διαγράμματα καθώς και σε πίνακες για κάθε παρόμοια τιμή του φίλτρου με το ποσό της συναθροιστικής τιμής της μέτρησης.

Στη δεύτερη αναφορά γίνεται σύγκριση ως προς την αναμενόμενη τιμή ενός φίλτρου και την πραγματική τιμή. Η αναμενόμενη τιμή υπολογίζεται για κάθε φίλτρο των διαστάσεων ως προς το ανώτερο επίπεδο και ως προς το σύνολο εγγραφών. Τα αποτελέσματα απεικονίζονται σε διαγράμματα καθώς και σε πίνακες, ώστε να γίνουν οι συγκρίσεις των αναμενόμενων τιμών και των πραγματικών.

Το σύστημα που υλοποιήθηκε ελέγχθηκε και αξιολογήθηκε πειραματικά. Για την εκάστοτε βάση δεδομένων που έγινε η ανάλυση δημιουργήθηκαν πειράματα, αλλά και οι αναμενόμενες “ορθές” ερωτήσεις χειρωνακτικά. Στη συνέχεια έγινε έλεγχος αν οι ερωτήσεις που δημιουργήθηκαν από το σύστημα είναι οι αναμενόμενες, καθώς και αν τα αποτελέσματά τους είναι ίδια με αυτά των “ορθών” ερωτήσεων που εκτελέστηκαν στη βάση δεδομένων. Επιπλέον, βελτιώθηκε ο αλγόριθμος σύγκρισης KPI με παρόμοιες τιμές. Αρχικά γινόταν εκτέλεση ερώτησης στη βάση δεδομένων για κάθε παρόμοια τιμή και βελτιώθηκε ως προς μια εκτέλεση ερώτησης με βάση της τιμής του ανωτέρου επιπέδου. Στα πειράματα που διεξήχθησαν απεικονίζονται οι χρόνοι εκτέλεσης συγκριτικά με τους δυο αλγορίθμους και παρατηρήθηκε πως η διαφορά μεταξύ τους είναι πολύ μεγάλη. Είναι σαφώς πιο γρήγορος ο αλγόριθμος που υλοποιεί μια εκτέλεση ερώτησης στη βάση δεδομένων, ώστε να ληφθούν τα αποτελέσματα παρόμοιων τιμών και των ποσών των μετρήσεων. Επιπλέον, στα πειράματα που πραγματοποιήθηκαν, παρατηρήθηκε σε όλους τους αλγορίθμους, πως κατά την κλιμάκωση των φίλτρων που διέθεταν οι KPI, ο χρόνος εκτέλεσης είχε την τάση να αυξάνεται.

6.2 Μελλοντικές επεκτάσεις

Σαν μελλοντικές επεκτάσεις στο σύστημα θα μπορούσαν να συμπεριληφθούν περισσότερες συγκρίσεις και προβλέψεις. Αρχικά, θα μπορούσαν να προστεθούν συγκρίσεις μεταξύ δύο δεικτών KPIs και τα αποτελέσματά τους να αναπαρασταθούν σε ένα διάγραμμα, ώστε η κάθε τιμή για κάθε φίλτρο από τους πίνακες διαστάσεων του ενός δείκτη να απεικονίζεται συγκριτικά με το αποτέλεσμα για την τιμή του φίλτρου από τον άλλο δείκτη. Στο σύστημα που υλοποιήθηκε έχει πραγματοποιηθεί πρόβλεψη της αναμενόμενης τιμής ως προς το ανώτερο επίπεδο, ως προς το γενικό σύνολο και σαν μελλοντική επέκταση θα μπορούσε να υλοποιηθεί και ως προς το χαμηλότερο επίπεδο από αυτό που έχει δηλώσει ο χρήστης σαν φίλτρο από τον πίνακα διαστάσεων, με αυτό τον τρόπο η αναμενόμενη τιμή θα παρέχεται από κάθε δυνατή πτυχή. Επιπλέον, άλλη μια

επέκταση θα ήταν, αν η αναμενόμενη τιμή που βρέθηκε είναι μεγαλύτερη από την πραγματική τιμή για μια παράμετρο του πίνακα διαστάσεων, τότε θα βρεθεί ποιο είναι το σύνολο τιμών από το χαμηλότερο επίπεδο, που επιφέρει αυτή την απόκλιση, και θα γίνεται αναπαράσταση του αποτελέσματος στη διεπαφή του χρήστη.

ΒΙΒΛΙΟΓΡΑΦΙΑ

- [Abe16] Mari Abe. Analyzing Business Processes by Automatically Detecting KPI Thresholds. 2016 IEEE International Conference on Services Computing, pp. 187-194.
- [AuSa17] Paulo Roberto Martins de Andrade, Dr. Samira Sadaoui. Improving Business Decision Making based on KPI Management System. 2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC), pp. 1280-1285.
- [Chart1] Chart.js is available under the MIT license. The manual of Charts available at <https://www.chartjs.org/docs/latest> Last accessed: 2022-05-25
- [Gat+22] Sneha Gathani, Madelon Hulsebos, James Gale, Peter J. Haas, Çagatay Demiralp. Augmenting Decision Making via Interactive What-If Analysis. CIDR'22, January 9-12, 2022, Chaminade CA, USA.
- [HIN+19] Andreas Hinderks, Martin Schreppb, Francisco José Domínguez Mayoa, María José Escalonaa, Jörg Thomaschewski. Developing a UX KPI based on the user experience questionnaire. Computer Standards & Interfaces 65 (2019), pp. 38-44.
- [LQKÖ21] Doris Jung-Lin Lee, Abdul Quamar, Eser Kandogan, Fatma Özcan. Boomerang: Proactive Insight-Based Recommendations for Guiding Conversational Data Analysis. SIGMOD '21, June 20–25, 2021.
- [MDHZ21] Pingchuan Ma, Rui Ding, Shi Han, Dongmei Zhang. MetaInsight: Automatic Discovery of Structured Knowledge for Exploratory Data Analysis. SIGMOD '21, June 20–25, 2021.
- [Micro2] Microsoft Corp. The manual of .Net Framework. Available at <https://docs.microsoft.com/en-us/dotnet/framework/get-started/overview> Last accessed: 2021-09-15
- [Micro3] Microsoft Corp. The manual of C#. Available at <https://docs.microsoft.com/en-us/dotnet/csharp/programming-guide/> Last accessed: 2022-03-31

- [Micro4] Microsoft Corp. The manual of MS SQL Server. Available at <https://docs.microsoft.com/en-us/sql/t-sql/language-reference?view=sql-server-ver15> Last accessed: 2022-01-11
- [Micro5] Microsoft Corp. The manual of ASPNET. Available at <https://docs.microsoft.com/en-us/aspnet/overview> Last accessed: 2022-05-13
- [Micro6] Microsoft Corp. The manual of WCF Service. Available at <https://docs.microsoft.com/en-us/dotnet/framework/wcf/whats-wcf> Last accessed: 2022-12-16
- [Micro7] Microsoft Corp. The manual of Class-Libraries. Available at <https://docs.microsoft.com/en-us/dotnet/standard/class-libraries> Last accessed: 2022-12-01
- [VAZI14] Alejandro Vaisman Esteban Zim'anyi Data Warehouse Systems Design and Implementation. Springer-Verlag Berlin Heidelberg 2014

ΣΥΝΤΟΜΟ ΒΙΟΓΡΑΦΙΚΟ

Η Δήμητρα Γεωργιάδου είναι μεταπτυχιακή φοιτήτρια στο Τμήμα Επιστήμης και Μηχανικής Υπολογιστών του Πανεπιστημίου Ιωαννίνων. Έλαβε το πτυχίο από το τμήμα Μηχανικών Ηλεκτρονικών Υπολογιστών και Πληροφορικής του Πανεπιστημίου Ιωαννίνων το 2017. Έκανε την πρακτική της στην εταιρία NATECH, όπου και εργάστηκε ως Μηχανικός Λογισμικού για 2 χρόνια και 8 μήνες. Από τον Απρίλιο του 2020 εργάζεται ως Μηχανικός Λογισμικού στην εταιρεία P&I.