

Σχεδιασμός Βάσεων Δεδομένων
Διδάσκων: Ιωάννης Κωτίδης

Εαρινό εξάμηνο 2020-2021

Δεύτερη Εργασία

Ανάθεση: **24-05-2021**

Παράδοση: **06-06-2021 Ώρα (23:55)**

Οδηγίες

- *Η εργασία είναι ατομική και υποχρεωτική.*
- *Η υποβολή της εργασίας πρέπει να γίνει στο eclass.*
- *Το παραδοτέο σας θα πρέπει να είναι ένα αρχείο PDF με όνομα AM.pdf (όπου AM είναι ο αριθμός μητρώου σας. π.χ. "3180001.pdf").*
- *Πιθανή αντιγραφή θα τιμωρείται με μηδενισμό όλων των εμπλεκομένων.*

Δημιουργία Αποθήκης Δεδομένων

Μία εταιρεία η οποία δραστηριοποιείται στον χώρο του τουρισμού διαθέτει έναν αριθμό κατασκηνώσεων (camping) στην ηπειρωτική ελληνική περιφέρεια και στα ελληνικά νησιά. Η εταιρεία συνεργάζεται με έναν σημαντικό αριθμό τουριστικών γραφείων (πελάτες της εταιρείας) από διάφορες ευρωπαϊκές χώρες. Η εταιρεία διαθέτει εφαρμογή μέσω της οποίας τα τουριστικά γραφεία μπορούν να κάνουν κρατήσεις. Τα στοιχεία των κρατήσεων καταγράφονται σε μία βάση δεδομένων.

Η διοίκηση της εταιρείας ενδιαφέρεται να αναπτύξει μια αποθήκη δεδομένων, ώστε να είναι σε θέση να αναλύει τα δεδομένα και να παράγει στατιστικές αναφορές σχετικά με την πορεία των κρατήσεων και τον αριθμό των ενοικιαστών. Οι απαιτήσεις της διοίκησης εστιάζουν στην ανάλυση των κρατήσεων ανά χώρα, τουριστικό γραφείο, κατασκήνωση, κατηγορία θέσης, καθώς και οποιονδήποτε συνδυασμό αυτών. Εξυπακούεται ότι στην ανάλυση των δεδομένων θα πρέπει να ληφθεί υπόψη ο παράγοντας του χρόνου έτσι ώστε, η διοίκηση της εταιρείας να είναι σε θέση να αναλύσει τα στοιχεία των κρατήσεων σε ημερήσια και μηνιαία βάση για το διάστημα ενός ή περισσότερων ετών.

Τα δεδομένα που θα τροφοδοτούν την αποθήκη είναι διαθέσιμα υπό την μορφή ενός γραμμογραφημένου αρχείου κειμένου (αρχείο CAMPDATA.TXT), το οποίο παράγεται από την εφαρμογή κρατήσεων της εταιρείας.

Καλείστε να σχεδιάσετε και να υλοποιήσετε την παραπάνω αποθήκη δεδομένων προκειμένου να αυξήσετε την αποτελεσματικότητα της διεξαγωγής χρήσιμων στατιστικών στοιχείων, μειώνοντας ταυτόχρονα τον χρόνο εκτέλεσης των επερωτήσεων. Στην συνέχεια να τροφοδοτήσετε την αποθήκη με τα δεδομένα του αρχείου "CAMPDATA.TXT" και να εκτελέσετε ορισμένες επερωτήσεις για την παραγωγή χρήσιμων στατιστικών αναφορών .

Περιγραφή Αρχείου CAMPDATA.TXT

Το αρχείο **CAMPDATA.TXT** περιέχει 1.311.107 εγγραφές. Κάθε εγγραφή αποτελείται από 16 πεδία τα οποία διαχωρίζονται με τον χαρακτήρα "|" (pipe). Ακολουθεί η περιγραφή των πεδίων του αρχείου CAMPDATA.TXT.

		CAMPDATA.TXT
custID	integer	Κωδικός Πελάτη. Πελάτες της εταιρείας είναι τα τουριστικά γραφεία με τα οποία συνεργάζεται.
fname	varchar(30)	Όνομα υπευθύνου του τουριστικού γραφείου.
lname	varchar(30)	Επώνυμο υπευθύνου του τουριστικού γραφείου.
clD	integer	Κωδικός Χώρας τουριστικού γραφείου
country	varchar(30)	Χώρα τουριστικού γραφείου
bookID	integer	Κωδικός Κράτησης
bookDate	date	Ημερομηνία υποβολής της κράτησης.
campCode	char(3)	Κωδικός Κατασκήνωσης
campName	Varchar(50)	Όνομα Κατασκήνωσης
empno	integer	Αριθμός θέσης (emplacement no). Κάθε κατασκήνωση διαθέτει ένα σύνολο αριθμημένων θέσεων.
catCode	char(1)	Κωδικός κατηγορίας θέσης
category	varchar(20)	Κατηγορία θέσης
unitCost	numeric(4,2)	Κόστος διανυκτέρευσης ανά άτομο.
startDate	date	Ημερομηνία πρώτης διανυκτέρευσης (έναρξη της κράτησης).
overnights	integer	Αριθμός διανυκτερεύσεων
persons	integer	Αριθμός ατόμων

Υπόδειξεις:

1. Για πληρέστερη κατανόηση των δεδομένων δείτε το υπόδειγμα της φόρμας κρατήσεων που υπάρχει στην τελευταία σελίδα.
2. Για την χρονική ανάλυση των κρατήσεων δεν μας ενδιαφέρει η ημερομηνία υποβολής της κράτησης (bookDate) αλλά η ημερομηνία έναρξης (startDate).

Ζήτημα Πρώτο [μονάδες 35]

Να δημιουργήσετε το λογικό σχήμα της αποθήκης δεδομένων και να το τροφοδοτήσετε με τα απαραίτητα δεδομένα. Συγκεκριμένα:

1. Να δημιουργήσετε μία βάση δεδομένων με όνομα **CAMPDW (Camping Data Warehouse)**. Στη συνέχεια να δημιουργήσετε τον πίνακα **campdata** στον οποίο να φορτώσετε τα δεδομένα του αρχείου **CAMPDATA.TXT** χρησιμοποιώντας την παρακάτω εντολή:

```
BULK INSERT campdata
FROM 'C:\data\CAMPDATA.TXT' !!! Προσαρμόστε το path
WITH (FIRSTROW =2, FIELDTERMINATOR='|', ROWTERMINATOR = '\n');
```
2. Να υλοποιήσετε το λογικό σχήμα της αποθήκης δεδομένων το οποίο θα πρέπει να έχει την μορφή αστέρα (Star Schema).
3. Να γράψετε κατάλληλες εντολές σε γλώσσα SQL, οι οποίες θα τροφοδοτούν το σχήμα της αποθήκης με τα απαραίτητα στοιχεία από τον πίνακα **campdata**.
4. Να αναπαραστήσετε διαγραμματικά το σχήμα της αποθήκης χρησιμοποιώντας την επιλογή "Database diagrams" του SQL Server Management Studio.

Η δημιουργία του λογικού σχήματος και η τροφοδότηση της αποθήκης με τα δεδομένα θα γίνουν με την εκτέλεση ενός SQL script το οποίο θα πρέπει να γράψετε.

Όλα τα παραπάνω (1,2,3,4) αποτελούν το παραδοτέο του πρώτου ζητήματος.

Ζήτημα Δεύτερο [μονάδες 35]

Χρησιμοποιώντας την αποθήκη δεδομένων που δημιουργήσατε στο προηγούμενο ζήτημα, να γράψετε και να εκτελέσετε επερωτήσεις σε γλώσσα SQL, οι οποίες να απαντούν στα ακόλουθα ερωτήματα (απαιτήσεις) της διοίκησης της εταιρείας:

1. Εμφανίστε έναν κατάλογο με τους πρώτους εκατό καλύτερους πελάτες της εταιρείας. Ο κατάλογος θα περιέχει την χώρα, το ονοματεπώνυμο του υπεύθυνου του τουριστικού γραφείου και την συνολική αξία των κρατήσεων που έχουν γίνει από το συγκεκριμένο τουριστικό γραφείο.
2. Εμφανίστε ένα κατάλογο με την συνολική αξία των κρατήσεων κάθε κατασκήνωσης ανά κατηγορία θέσης για το έτος 2000 (Όνομα Κατασκήνωσης, Κατηγορία Θέσης, Αξία κρατήσεων).
3. Εμφανίστε έναν κατάλογο με την συνολική αξία των κρατήσεων ανά κατασκήνωση σε μηνιαία βάση για το έτος 2018.
4. Η διοίκηση της εταιρείας θέλει μία αναφορά που θα περιέχει τις ακόλουθες πληροφορίες.
 - a. Τον συνολικό αριθμό των ενοικιαστών
 - b. Τον αριθμό των ενοικιαστών ανά έτος
 - c. Τον αριθμό των ενοικιαστών ανά έτος και κατασκήνωση
 - d. Τον αριθμό των ενοικιαστών ανά έτος, κατασκήνωση και κατηγορίας θέσης.Γράψτε μια επερώτηση σε γλώσσα SQL η οποία να παράγει την παραπάνω αναφορά.
5. Εμφανίστε τα ονόματα των κατασκηνώσεων που δέχθηκαν περισσότερους ενοικιαστές το έτος 2018 σε σχέση με το 2017. Για το συγκεκριμένο ερώτημα επιτρέπεται η δημιουργία όψεων.

Ζήτημα τρίτο [μονάδες 15]

Γράψτε μια επερώτηση σε γλώσσα SQL το αποτέλεσμα της οποίας είναι η δημιουργία ενός κύβου που αναλύει κάποια μέτρηση με βάση τρεις διαστάσεις.

- a. Εξηγήστε τι περιέχει το κάθε κελί του κύβου.
- b. Γράψτε όλα τα group by που περιέχει ο κύβος.

Ζήτημα Τέταρτο [μονάδες 15]

Το ζήτημα αυτό είναι θεωρητικό. Ο SQL SERVER δεν υποστηρίζει bitmap ευρετήρια. Για να απαντήσετε βασιστείτε στις διαλέξεις του μαθήματος.

Έστω η διάσταση **countries (cid, country)** που περιέχει τις χώρες με τις οποίες συνεργάζεται η εταιρεία και ο πίνακας fact με τα στοιχεία των κρατήσεων. Έχουμε δημιουργήσει το παρακάτω ευρετήριο:

```
CREATE BITMAP INDEX fact_countries_idx
ON fact (countries.country)
FROM fact, countries
WHERE fact.cid=countries.cid
```

Θεωρείστε επίσης το ακόλουθο ερώτημα και ένα μικρό δείγμα (8 εγγραφές) του αποτελέσματος.

```
select row_number() over (order by campCode,empno,catCode)
as Row#, campCode,empno,catCode,
countries.country, fact.cost
from countries,fact
where countries.cid = fact.cid
```

row#	campCode	empno	catCode	country	cost
69	ELB	62	C	Spain	15.00
70	KIS	9	A	Spain	50.00
71	ROS	191	C	Denmark	60.00
72	APL	55	C	Austria	120.00
73	ROS	87	C	Austria	90.00
74	DIS	75	C	Germany	75.00
75	APL	42	C	Germany	30.00
76	DIS	83	C	Germany	240.00

Δώστε την μορφή και το περιεχόμενο του ευρετηρίου για τις παραπάνω εγγραφές. Να λάβετε υπόψη όλες τις χώρες της διάστασης countries.

ΥΠΟΔΕΙΓΜΑ ΦΟΡΜΑΣ ΚΡΑΤΗΣΕΩΝ

<p>Αριθμός Κράτησης: 150 Ημερομηνία: 19 Μαΐου 2019</p> <p>Στοιχεία υπεύθυνου τουριστικού γραφείου</p> <p>Επώνυμο : Smith Όνομα : Richard Κωδικός: 1001</p>										
Κωδικός Κατασκήνωσης	Όνομα Κατασκήνωσης	Αριθμός Θέσης	Κωδικός Κατηγορίας Θέσης	Κατηγορία Θέσης	Έκταση Θέσης M ²	Κόστος Θέσης	Ημερομηνία Έναρξης	Διανυκτερεύσεις	Αριθμός Ατόμων	Κόστος (Ευρώ)
APL	Apollonia	3	A	Tent	10	5	12/06/2019	3	2	30
APL	Apollonia	28	C	Camper Van	30	15	12/06/2019	4	3	180
DIS	Dionysus	3	B	Caravan	20	10	12/06/2019	3	2	60
DIS	Dionysus	15	A	Tent	10	5	15/06/2019	1	4	20
										ΣΥΝΟΛΙΚΟ ΚΟΣΤΟΣ: 290.00

Υπόδειξη: Το κόστος κάθε γραμμής προκύπτει ως εξής: Κόστος Θέσης * Διανυκτερεύσεις * Αριθμός ατόμων