



ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ

ΣΥΣΤΗΜΑΤΑ ΔΙΑΧΕΙΡΙΣΗΣ ΚΑΙ ΑΝΑΛΥΣΗΣ ΔΕΔΟΜΕΝΩΝ

2^η Εργασία

Όνοματεπώνυμο:

Μαρία Σχοινάκη

Αριθμός Μητρώου:

3210191

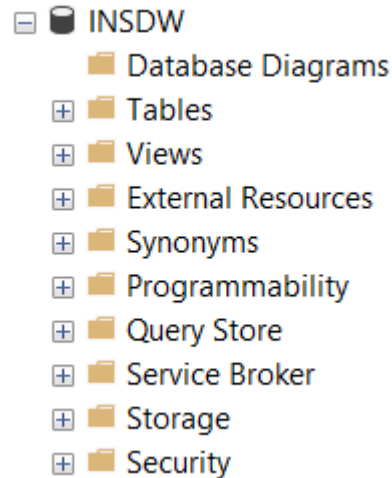
Email:

p3210191@aueb.gr

Ζήτημα 1^ο

1.

a. Δημιουργούμε την Βάση Δεδομένων με όνομα **INSDW**:



b. Δημιουργούμε τον πίνακα **inspections_data**:

```
CREATE TABLE inspections_data (
    rid INT,
    lat FLOAT,
    lon FLOAT,
    insdate DATE,
    insyear INT,
    insmonth INT,
    insday INT,
    insweekday INT,
    inscode INT,
    instype NVARCHAR(100),
    criticalIssue INT,
    nonCriticalIssue INT,
    vcode INT,
    vdescription NVARCHAR(255),
    vcategory NVARCHAR(255)
);
```

c. Φορτώνουμε τα δεδομένα στον πίνακα **inspections_data**:

```
-- Φόρτωση δεδομένων στον πίνακα inspections_data
BULK INSERT inspections_data
FROM 'C:\Users\shina\Downloads\inspections_data\inspections_data.txt'
WITH (
    DATAFILETYPE = 'widechar',
    FIRSTROW = 2,
    FIELDTERMINATOR = '|',
    ROWTERMINATOR = '\n'
);
```

2. Από την περιγραφή βρίσκουμε τις εξής **διαστάσεις** (dimensions):

*Η υπηρεσία δημόσιας υγιεινής ενδιαφέρεται να αναπτύξει μια αποθήκη για την άντληση χρήσιμων πληροφοριών σχετικά με τα στοιχεία των επιθεωρήσεων. Οι απαιτήσεις της υπηρεσίας δημόσιας υγιεινής εστιάζουν μεταξύ άλλων στην ανάλυση του αριθμού των επιθεωρήσεων και των παραβάσεων ανά **τύπο επιθεώρησης**, **κατηγορία παράβασης**, **περιοχή**, καθώς και οποιονδήποτε συνδυασμό αυτών. Εξυπακούεται ότι στην ανάλυση των δεδομένων θα πρέπει να ληφθεί υπόψη και ο παράγοντας του **χρόνου** έτσι ώστε, η υπηρεσία να είναι σε θέση να παράγει στατιστικές αναφορές στοιχεία από τα ευρήματα των επιθεωρήσεων ανά έτος, μήνα, ημέρα κ.λπ.*

Το **star schema** θα περιλαμβάνει έναν πίνακα για κάθε **διάσταση** (dimension). Η δημιουργία των πινάκων αυτών φαίνεται παρακάτω.

ΤΥΠΟΣ ΕΠΙΘΕΩΡΗΣΗΣ

```
CREATE TABLE inspection_type (
    inscode INT PRIMARY KEY,
    instype NVARCHAR(100),
);
```

ΚΑΤΗΓΟΡΙΑ ΠΑΡΑΒΑΣΗΣ

```
CREATE TABLE violation (
    vcode INT PRIMARY KEY,
    vdescription NVARCHAR(255),
    vcategory NVARCHAR(255)
);
```

ΠΕΡΙΟΧΗ

```
CREATE TABLE area (
    rid INT PRIMARY KEY,
    lat FLOAT,
    lon FLOAT,
);
```

ΧΡΟΝΟΣ

```
CREATE TABLE timeinfo (
    time_key DATE PRIMARY KEY,
    insyear INT,
    insmonth INT,
    insday INT,
    insweekday INT,
);
```

Project 2

Τέλος δημιουργείται ο πίνακας “fact table”, ο οποίος περιέχει τα γεγονότα, δηλαδή επιθεωρήσεις. Ο fact table περιέχει τα κλειδιά των **διαστάσεων** (dimension) ως ξένα κλειδιά, για αυτό και τον δημιουργούμε τελευταίο.

```
CREATE TABLE inspection (  
    rid INT,  
    inscode INT,  
    vcode INT,  
    time_key DATE,  
    criticalIssue INT,  
    nonCriticalIssue INT,  
    PRIMARY KEY (rid, inscode, vcode, time_key),  
    FOREIGN KEY (rid) REFERENCES area(rid),  
    FOREIGN KEY (inscode) REFERENCES inspection_type(inscode),  
    FOREIGN KEY (vcode) REFERENCES violation(vcode),  
    FOREIGN KEY (time_key) REFERENCES timeinfo(time_key)  
);
```

Από την περιγραφή εντοπίζουμε και κάποια μετρήσιμα τα οποία θα ήταν πλεονασμός να μπουν σαν ξεχωριστά πεδία, καθώς υπολογίζονται πολύ εύκολα, δυναμικά με την χρήση της count()

3. Τροφοδοτούμε τους πίνακες με δεδομένα από τον πίνακα **inspections_data**:

```
insert into inspection_type  
select distinct inscode, instype  
from inspections_data
```

```
insert into violation  
select distinct vcode, vdescription, vcategory  
from inspections_data
```

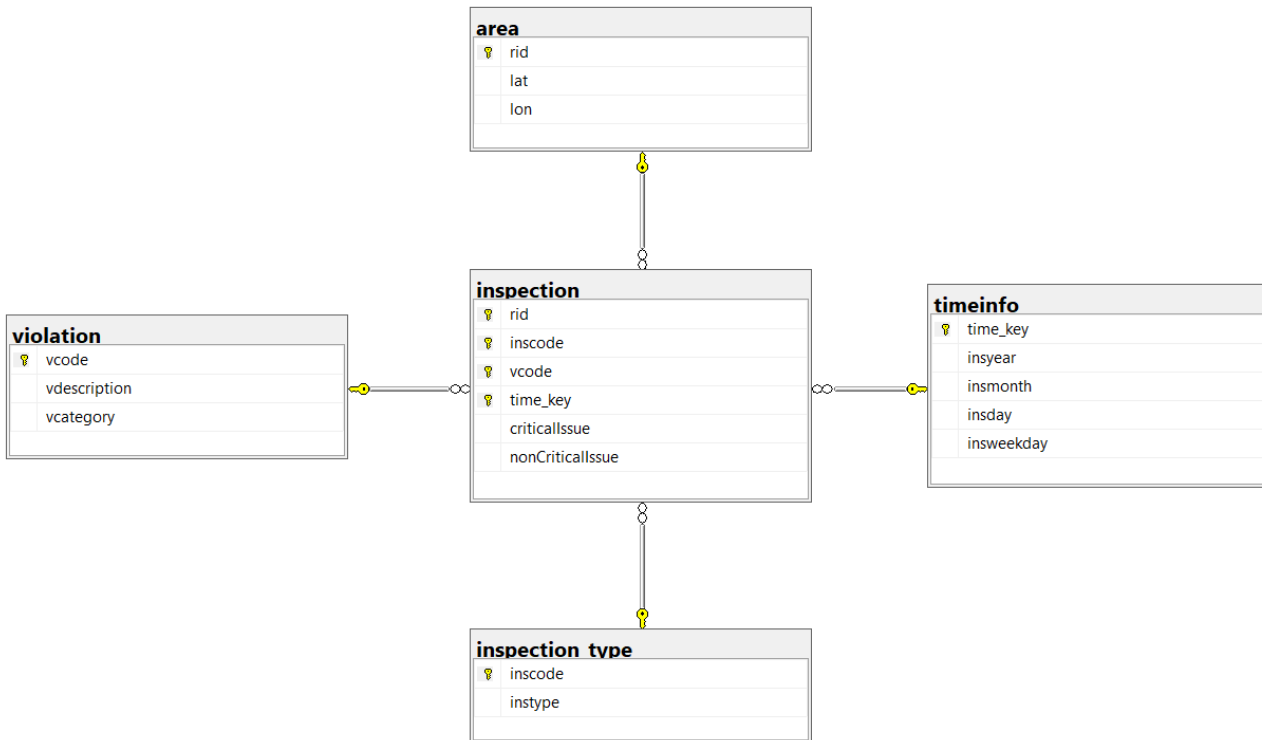
```
insert into area  
select distinct rid, lat, lon  
from inspections_data
```

```
insert into timeinfo  
select distinct insdate, insyear, insmonth, insday, insweekday  
from inspections_data
```

```
insert into inspection  
select rid, inscode, vcode, insdate, criticalIssue, nonCriticalIssue  
from inspections_data
```

Project 2

4. Διαγραμματική αναπαράσταση του σχήματος αστέρα της αποθήκης δεδομένων:



Ζήτημα 2°

1. Εμφανίστε έναν κατάλογο με τον αριθμό των επιθεωρήσεων ανά έτος και τύπο επιθεώρησης (instype). Ο κατάλογος πρέπει να είναι ταξινομημένος με βάση το έτος σε φθίνουσα διάταξη.

```
select insyear as 'Year', instype as 'Type', count(*) as 'Number of inspections'
from timeinfo, inspection_type, inspection
where timeinfo.time_key = inspection.time_key and inspection_type.inscode =
inspection.inscode
group by insyear, instype
order by insyear desc
```

2. Εμφανίστε έναν κατάλογο με τα παρακάτω στοιχεία για κάθε εστιατόριο:

- Κωδικός εστιατορίου
- Συντεταγμένες
- Αριθμός μη κρίσιμων ζητημάτων που εντοπίστηκαν συνολικά από όλες τις επιθεωρήσεις που έγιναν στο συγκεκριμένο εστιατόριο.
- Αριθμός κρίσιμων ζητημάτων που εντοπίστηκαν συνολικά από όλες τις επιθεωρήσεις που έγιναν στο συγκεκριμένο εστιατόριο.
- Αριθμός ζητημάτων (κρίσιμων και μη κρίσιμων) που εντοπίστηκαν συνολικά από όλες τις επιθεωρήσεις που έγιναν στο συγκεκριμένο εστιατόριο.

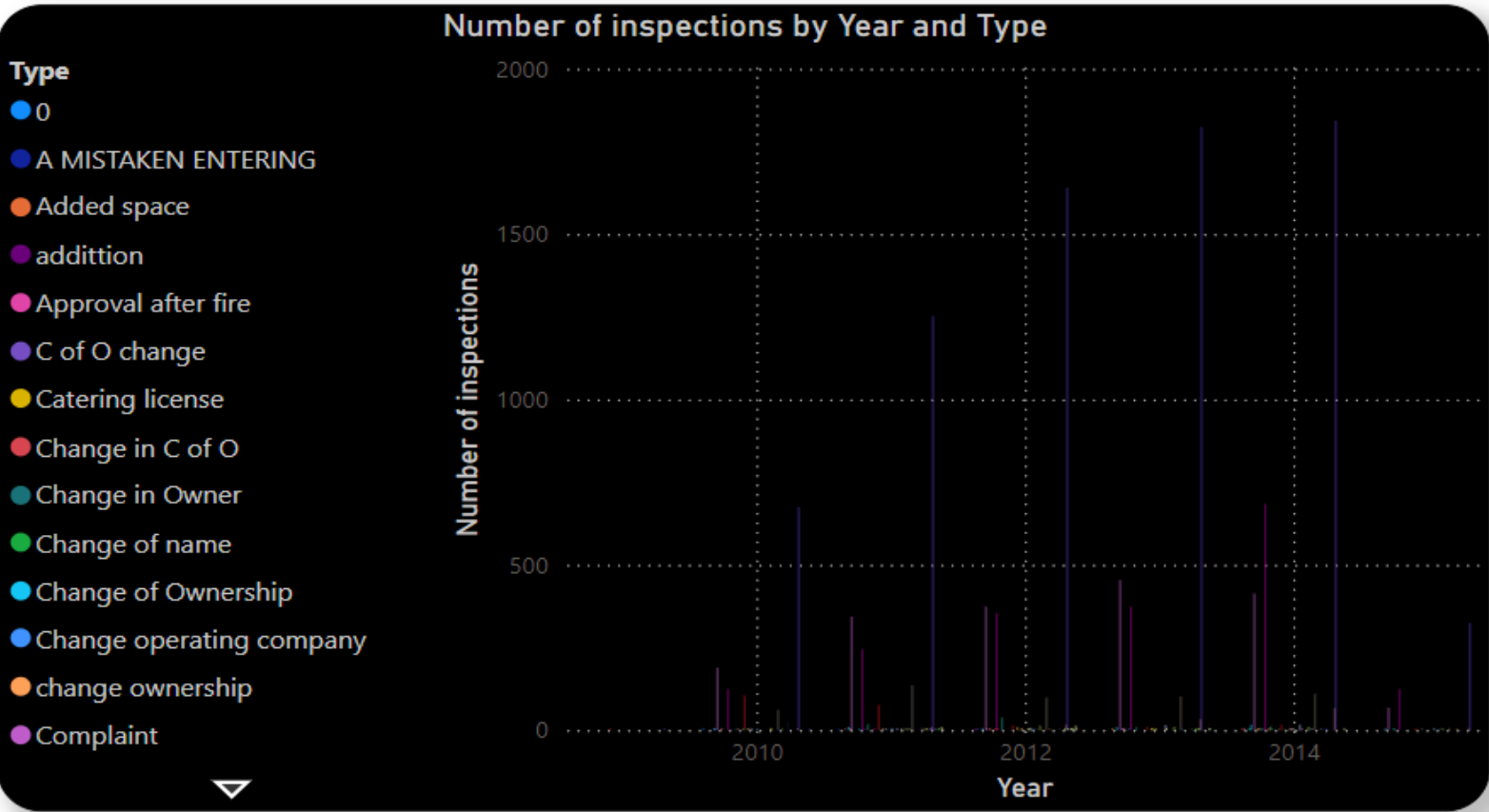
Ο κατάλογος να είναι ταξινομημένος σε φθίνουσα διάταξη βάσει του αριθμού ζητημάτων (κρίσιμων και μη κρίσιμων).

```
select area.rid as 'Restaurant Code', lat as 'Latitude', lon as 'Longitude',
sum(nonCriticalIssue) as 'Non critical issues',
sum(criticalIssue) as 'Critical issues', sum(nonCriticalIssue + criticalIssue) as
'Total issues'
from area, inspection
where area.rid = inspection.rid
group by area.rid, lat, lon
order by 'Total issues' desc
```

3. Γράψτε μια επερώτηση σε γλώσσα SQL το αποτέλεσμα της οποίας είναι η δημιουργία ενός κύβου (data cube), κάθε κελί του οποίου περιέχει τον συνολικό αριθμό των κρίσιμων ζητημάτων που εντοπίστηκαν από όλες τις επιθεωρήσεις ανά τύπο επιθεώρησης (instype), κατηγορία παράβασης (vcategory) και έτος επιθεώρησης (insyear).

```
select instype AS 'Type', vcategory as 'Category', insyear as 'Year',
sum(criticalIssue) as 'Critical issues'
from inspection_type, violation, timeinfo, inspection
where inspection_type.inscode = inspection.inscode and violation.vcode =
inspection.vcode and timeinfo.time_key = inspection.time_key
group by cube (instype, vcategory, insyear)
```

Ζήτημα 3ο



| Resaturant Code | Latitude | Longitude |
|-----------------|----------|-----------|
| 757 | 38,93 | -76,98 |
| 862 | 38,91 | -77,02 |
| 916 | 38,89 | -76,95 |
| 976 | 38,94 | -76,99 |
| 1024 | 38,92 | -77,00 |
| 1826 | 38,91 | -77,02 |
| 2973 | 38,91 | -77,05 |
| 3470 | 38,90 | -77,03 |
| 3507 | 38,90 | -77,05 |
| 3510 | 38,90 | -77,03 |
| 3706 | 38,90 | -77,02 |
| 4288 | 38,88 | -77,00 |
| 4527 | 38,83 | -77,01 |
| 4599 | 38,89 | -77,02 |
| 4603 | 38,89 | -77,00 |
| 4692 | 38,94 | -77,02 |
| 5167 | 38,91 | -77,07 |
| 5580 | 38,92 | -77,04 |
| 47238 | 38,89 | -77,01 |
| 147691 | 38,91 | -77,04 |