

Student : Agiopoulos S. Kaptsikas (akapsikas@aueb.gr, f3352208)

M.Sc. in Data Science
School of Information Sciences
Athens University of Economics and Business
Full Time, Fall, 2022

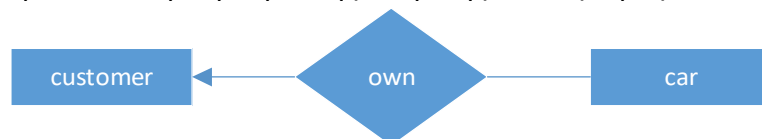
Introduction to Data Management and Engineering
Second Assignment: E-R Diagrams and Database Programming with Python

1. Αρχικά, λαμβάνοντας υπόψιν την δοθείσα επιχειρηματική περίπτωση, άντλησα πληροφορίες σχετικά με τις ανάγκες δεδομένων, που θα έπρεπε να ικανοποιεί η βάση δεδομένων. Για παράδειγμα, κάποια θεμελιώδη ευρήματα αυτής της διαδικασίας ήταν τα εξής:

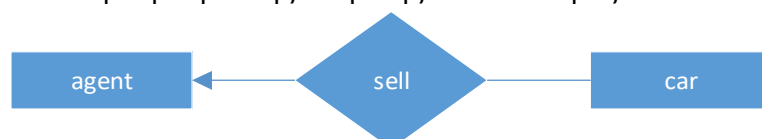
- a. Ένα συγκεκριμένο μοντέλο, μιας συγκεκριμένης μάρκας αυτοκινήτων της εταιρείας, μπορεί να προσφερθεί στους πελάτες της μέσω διάφορων επιλογών (λ.χ., μετρητά, δάνειο, μίσθωση, κ.λπ.), αλλά κάθε μεμονωμένο αυτοκίνητο δεν μπορεί παρά να προσφερθεί μόνο με ένα υποσύνολο (συμπεριλαμβανομένου του κενού) αυτών. Εδώ, παρατηρούμε ξεκάθαρα μια σχέση μεταξύ αυτοκινήτων και προσφορών τους, που επιβάλλει έναν ανά μεταξύ τους περιορισμό της λογικής «πολλά-προς-πολλά».



- b. Ένα συγκεκριμένο αυτοκίνητο, συγκεκριμένου μοντέλου και μάρκας, δεν μπορεί παρά να έχει αγοραστεί από έναν και μόνον πελάτη, ή να μην έχει αγοραστεί, δηλαδή να υπάρχει στην αποθήκη της εταιρείας διαθέσιμο προς πώληση. Εδώ, επίσης, παρατηρούμε μια λογική σχέση μεταξύ αυτοκινήτων και πελατών, που επιβάλλει περιορισμό της λογικής «ένας-προς-πολλά».



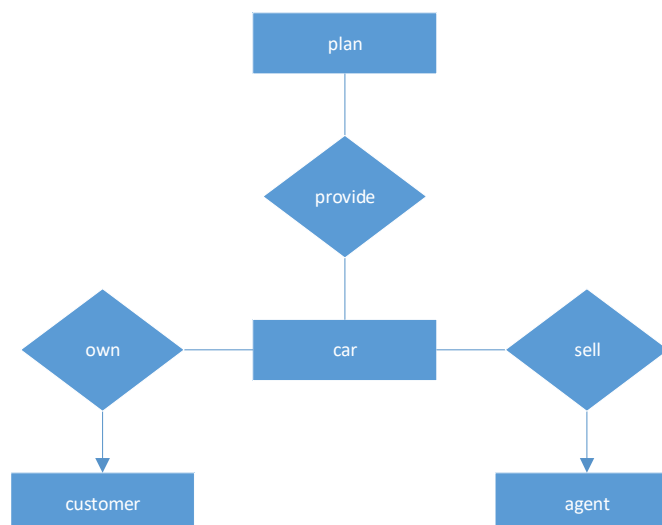
- c. Ένα συγκεκριμένο αυτοκίνητο, συγκεκριμένου μοντέλου και μάρκας αυτοκινήτων της εταιρείας, μπορεί να πωληθεί από έναν και μόνον αντιπρόσωπο της εταιρείας (ωστόσο, αυτό δεν σημαίνει ότι, γενικά, συγκεκριμένα μοντέλα, συγκεκριμένων μαρκών αυτοκινήτων, δεν μπορούν να πωληθούν από πολλούς αντιπροσώπους της εταιρείας, ταυτόχρονα). Εδώ, παρατηρούμε μια λογική σχέση μεταξύ αυτοκινήτων και αντιπροσώπων, που επιβάλλει περιορισμό της λογικής «πολλά-προς-έναν».



Όπως μπορούμε να κατανοήσουμε, με τη προηγούμενη σχεσιακή ανάλυση, ορίσαμε τέσσερεις διαφορετικές οντότητες του μοντέλου δεδομένων. Αυτές είναι οι εξής:

- i. Η οντότητα “car”, στην οποία θα καταγραφούν όλα τα αυτοκίνητα, που είτε έχουν πωληθεί σε πελάτες από τους αντιπροσώπους, είτε δεν έχουν πωληθεί ακόμα και παραμένουν στις αποθήκες των αντίστοιχων αντιπροσώπων.
- ii. Η οντότητα “plan”, στην οποία θα καταγραφούν οι διαφορετικές χρηματοδοτικές επιλογές προσφοράς αυτοκινήτων στους πελάτες από τους αντιπροσώπους.
- iii. Η οντότητα “agent”, στην οποία θα καταγραφούν οι αντιπρόσωποι, που συνεργάζονται με την εταιρεία.
- iv. Η οντότητα “customer”, στην οποία θα καταγραφούν οι πελάτες της εταιρείας, που έχουν αγοράσει αυτοκίνητα.

Σεβόμενοι αυτούς τους σχεσιακούς περιορισμούς, προερχόμενους από τη φύση του προβλήματος, μπορούμε να κατασκευάσουμε το παρακάτω σχεσιακό διάγραμμα οντοτήτων:



Αξίζει να σημειώσουμε, ότι τα δυαδικά σχεσιακά σύνολα οντοτήτων (binary relational entity sets) “customer ← own – car” και “agent ← sell – car”, δεν απαιτούν την δημιουργία των τρίτων, βοηθητικών οντοτήτων “own” και “sell”, καθώς οι σχέσεις αντιστοίχισής τους είναι του τύπου «ένα-προς-πολλά». Από την άλλη πλευρά, η σχέση “plan ← provide – car”, όντας της λογικής «πολλά-προς-πολλά», απαιτεί την δημιουργία της τρίτης, βοηθητικής οντότητας “provide”, η οποία θα αποθηκεύσει τα πρωτεύοντα κλειδιά των οντοτήτων “plan” και “car”, έτσι ώστε να επιτευχθούν αποτελεσματικά τυχούσες συνενώσεις τους.

2. Δύο βασικά ερωτήματα, των οποίων η εκτέλεση ενδέχεται να ζητείται συχνά από τους χρήστες της βάσης δεδομένων, είναι τα εξής:

- i. Πόσα αυτοκίνητα από κάθε συγκεκριμένο μοντέλο, συγκεκριμένης μάρκας, βρίσκονται διαθέσιμα προς πώληση στις αποθήκες της εταιρείας, για έναν μεμονωμένο αντιπρόσωπο; Αυτό το ερώτημα, θα βοηθήσει το προσωπικό των αντιπροσώπων, αφενός να ελέγχει την διαθεσιμότητα των αυτοκινήτων που προσφέρει, αλλά αφετέρου να είναι σε θέση να παραγγείλει μοντέλα αυτοκινήτων, που είναι σε έλλειψη.
- ii. Πόσο είναι το συνολικό ύψος των χρηματικών εισπρακτέων ποσών, που απομένουν να αποπληρωθούν από τους πελάτες, για κάθε είδος χρηματοδοτικής επιλογής προσφοράς αυτοκινήτων από την εταιρεία αυτοκινήτων, για έναν συγκεκριμένο αντιπρόσωπο; Υποθέστε ότι η εταιρεία έχει αναλάβει τον χρηματοδοτικό διακανονισμό των αγορών των πελατών της, οπότε έχει αναλάβει και την είσπραξη των αρχικών κεφαλαίων, βεβαρυμμένων των τόκων εισπρακτέων.

Δεδομένου, ότι όλες οι σχετικές πληροφορίες με τα δύο παραπάνω ερωτήματα βρίσκονται αποθηκευμένες στην οντότητα “car”, θα είχε νόημα να εφαρμοστεί ένα ευρετήριο πάνω σε αυτόν τον πίνακα, και συγκεκριμένα στις χαρακτηριστικές στήλες, που παρέχουν πληροφορίες για την ταυτότητα των αντιπροσώπων και των πελατών τους, καθώς με βάση αυτές κάθε αντιπρόσωπος θα επιλέξει εκείνα τα αυτοκίνητα, που είτε έχουν αγοραστεί από τους δικούς του πελάτες, είτε δεν έχουν αγοραστεί ακόμα, και παραμένουν διαθέσιμα στις αποθήκες του. Παρατηρήστε, ότι σε αυτά τα ερωτήματα εμπλέκονται και άλλες οντότητες/πίνακες, αλλά συνενώνονται με χρήση των πρωτευόντων κλειδιών τους. Ωστόσο, για αυτούς τους πίνακες, έχουν ήδη αυτόματα εφαρμοστεί ευρετήρια στα πρωτεύοντα κλειδιά τους από το πρόγραμμα της βάσης δεδομένων, και δεν χρειάζεται να κάνουμε καμία περαιτέρω δημιουργία άλλων ευρετηρίων.

3. Αν και γι’ αυτό το ερώτημα έχει δημιουργηθεί ξεχωριστό αρχείο Jupyter Notebook, στο οποίο περιέχεται, τόσο ο κώδικας, όσο και οι επεξηγήσεις του, εδώ, θα αναφερθούμε σε ορισμένες παρατηρήσεις που αφορούν στην γενικότερη λειτουργία του μενού επιλογών.

Γενικά, η λειτουργία του βασικού μενού της βάσης δεδομένων έχει επιδιωχτεί να προσφέρει τη μεγαλύτερη δυνατή ελευθερία επιλογών προς το χρήστη. Αυτό, σημαίνει ότι ο χρήστης δεν χρειάζεται να περιοριστεί σε ένα σημαντικά μικρότερο υποσύνολο δυνατοτήτων της βάσης δεδομένων, έτσι ώστε το μενού να λειτουργεί ομαλά.

Συγκεκριμένα, μόνο για τις επιλογές «Διαχείριση» και «Αναζήτηση», ο χρήστης μπορεί να κάνει ό,τι θα έκανε κάποιος που είχε άμεση πρόσβαση στη βάση δεδομένων. Όσον αφορά την επιλογή «Εκτέλεση συγκεκριμένης ερώτησης», οι ελευθερίες του χρήστη περιορίζονται ραγδαία (ορθώς, ως προς την εκφώνηση της άσκησης), καθώς μπορεί να εκτελέσει μόνον

προκαθορισμένα ερωτήματα, δίνοντας απλά τις απαραίτητες πληροφορίες, που θα κάνει την εκτέλεσή τους προσωποποιημένη (π.χ. εκτέλεση ερωτημάτων του ερωτήματος 2, συγκεκριμένα για έναν μεμονωμένο αντιπρόσωπο). Γενικά, όταν ο χρήστης ζητείται να παραχωρήσει πληροφορίες, θεωρείται δεδομένο ότι ξέρει τι πρέπει να εισαγάγει, και για αυτό το λόγο παραλείπονται πληροφορίες ως προς τους περιορισμούς του κάθε πεδίου. Θεωρητικά, και υπό πραγματικές συνθήκες, η χρήση του συγκεκριμένου προγράμματος θα προηγούνταν κάποιας εκπαίδευσης, στην οποία θα υποβάλλονταν το προσωπικό των αντιπροσώπων της εταιρείας.

Προχωρώντας σε μερικές τεχνικές λεπτομέρειες, είναι σημαντικό να σημειωθεί ότι ως «καλή» πρακτική χρήσης του μενού για τους χρήστες, είναι ο σεβασμός των περιορισμών των πεδίων των πινάκων, για τα οποία καλούνται οι χρήστες να παρέχουν πληροφορίες. Ειδικότερα, παρακάτω δίδονται περιπτώσεις, όπου ο χρήστης αλληλοεπιδρά με τη λειτουργία του μενού, και οι ενδεδειγμένοι τρόποι χρήσης:

- i. Αρχικά, όταν ο χρήστης καλείται να πλοηγηθεί εντός του μενού, πρέπει να ακολουθεί τους κάτωθι κανόνες, οι οποίοι ισχύουν γενικά για όλες τις περιπτώσεις.

User prompt:

choose action from:

- 1) **manage**, (could also be **insert**, ...
- 2) **search**, or **delete**, or **update**, ...
- 3) **query** or **Q1**, **Q2**, ...)

Expected Answer:

query, insert, delete, update, Q1, Q2

What the menu **does not** expect as an answer:

QuErY, q, 1., 1, 1), ...

- ii. Έπειτα, όταν ο χρήστης έχει πλοηγηθεί εντός του μενού, και καλείται να εισαγάγει τιμές σε πεδία κάποιου πίνακα, που έχει επιλέξει, πρέπει να ακολουθεί τους κάτωθι κανόνες. Υποθέστε, ότι ο χρήστης ζητείται να εισάγει μια *συνθήκη* και μια ενημερωμένη τιμή, για το πεδίο “vin”. Για τη πρώτη περίπτωση, *πρέπει* να συμπεριληφθούν και οι αντίστοιχοι λογικοί τελεστές, μαζί με την ενδεδειγμένη τιμή σύγκρισης, δηλαδή εκείνη που σέβεται το τύπο του πεδίου (εδώ INT). Στη δεύτερη περίπτωση, που ο χρήστης καλείται να εισαγάγει νέες τιμές σε υπάρχοντα πεδία, θα *πρέπει* αν προηγείται ο τελεστής της ανάθεσης.

```
enter condition for vin: ==, >, <, >=, <= 99
enter updated value for vin: =99
```

- iii. Συγκεκριμένα, για τη περίπτωση εισαγωγής δεδομένων σε πίνακα, ο χρήστης δεν χρειάζεται να συμπεριλάβει τον

τελεστή ανάθεσης, ούτε να σεβαστεί τον τύπο του πεδίου προς συμπλήρωση. Απλώς, εισάγει την τιμή του πεδίου:

enter value for vin: 99