

Εθνικό Μετσόβιο Πολυτεχνείο

Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών

**Εργαστήριο Βάσεων Δεδομένων**

**Εξαμηνιαία Εργασία:**

**Ανάπτυξη Βάσης Δεδομένων για Διαγωνισμό Μαγειρικής**

Ομάδα Project: 102

Αναγνώστου Νικόλαος: el21818

Μάκρας Ηλίας: el21207

Μαρκουλιδάκης Γεώργιος: el21050

Περιεχόμενα

[**1.** **Σχεδίαση της Βάσης** 2](#_Toc167635786)

[**1.1** **Εφαρμογές που χρησιμοποιήθηκαν** 2](#_Toc167635787)

[**1.2** **Διάγραμμα ER** 3](#_Toc167635788)

[**1.3** **Σχεσιακό Διάγραμμα** 5](#_Toc167635789)

[**2.** **Υλοποίηση της Βάσης** 6](#_Toc167635790)

[**2.1 DDL Script** 6](#_Toc167635791)

[**2.2 DML Script** 6](#_Toc167635792)

# **Σχεδίαση της Βάσης**

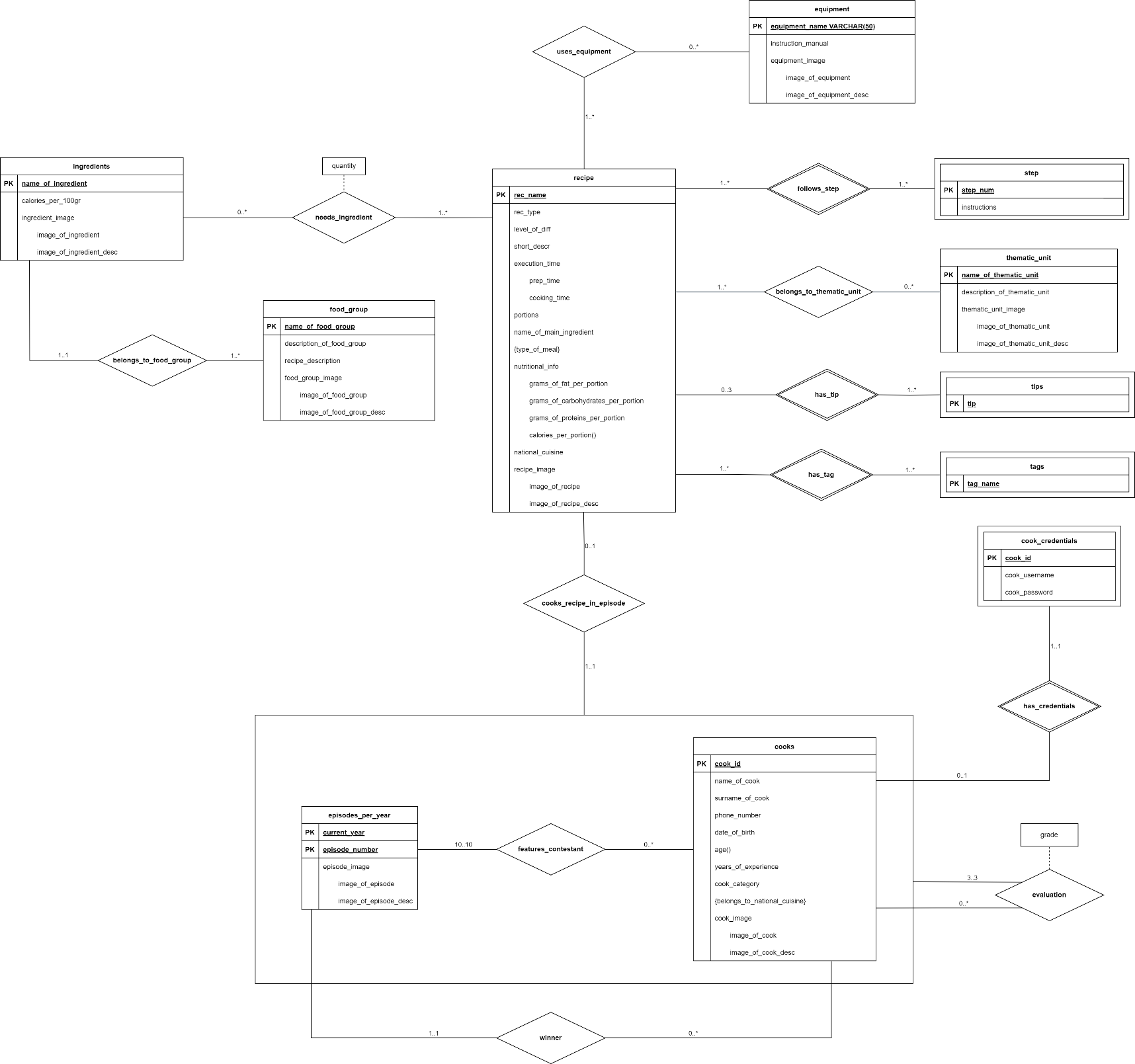
## **Εφαρμογές που χρησιμοποιήθηκαν**

Η εργασία μας σχεδιάστηκε και υλοποιήθηκε με τη χρήση της MySQL, μέσω του client MySQL Workbench για την εκτέλεση των SQL scripts που δημιουργούν τη βάση και εκτελούν την κλήρωση του διαγωνισμού, καθώς και των queries.

Τα δεδομένα που εισήχθησαν στην βάση δημιουργήθηκαν με κώδικα Python που περιλαμβάνει τις βιβλιοθήκες faker και faker\_food, οι οποίες λειτούργησαν ως γεννήτριες dummy data.

Όλα τα μέρη της εργασίας και οι υλοποιήσεις τους βρίσκονται στο σύνδεσμο για το GitHub repository: [*https://github.com/nikosanag/DB\_Project*](https://github.com/nikosanag/DB_Project)

## **Διάγραμμα ER**

Για την ορθή σχεδίαση της βάσης μας δημιουργήσαμε την αναπαράσταση των δεδομένων σε μοντέλο οντοτήτων-συσχετίσεων (ER). Το μοντέλο που σχεδιάσαμε φαίνεται στην ακόλουθη εικόνα:

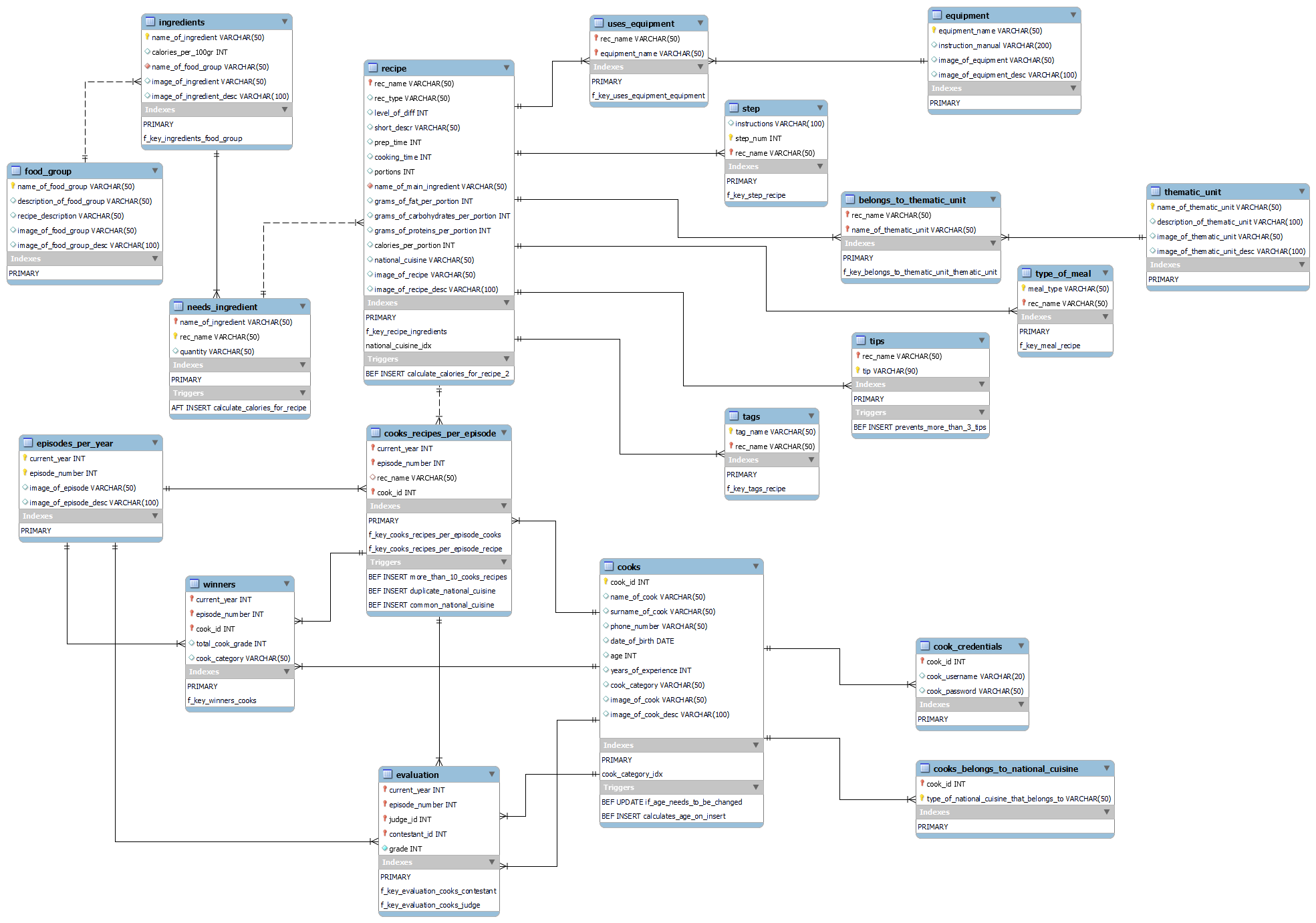
Η εικόνα φαίνεται καλύτερα στο αντίστοιχο αρχείο στο GitHub Repository της εργασίας με όνομα αρχείου *ER-Diagram.png*.

Στο παραπάνω διάγραμμα έχουμε επιλέξει να χρησιμοποιήσουμε σε όλες τις σχέσεις τους συμβολισμούς άνω και κάτω ορίου, για λόγους συνέπειας. Για τη δημιουργία του διαγράμματος έχουμε λάβει τις εξής παραδοχές:

* Οι οντότητες «thematic\_unit», «equipment» και «ingredients» είναι πιθανό να μη συσχετίζονται με κάποια συνταγή.
* Οι μάγειρες δεν σχετίζονται άμεσα με τις συνταγές που μπορούν να εκτελέσουν, αλλά μπορούν να εκτελέσουν οποιαδήποτε συνταγή ανήκει στις εθνικές κουζίνες στις οποίες είναι εξειδικευμένοι. Επειδή η εθνική κουζίνα δεν αποτελεί κάποια οντότητα στη μοντελοποίησή μας, η πληροφορία αυτή δεν φαίνεται στο διάγραμμα ER.
* Όπως θα αναφερθεί και παρακάτω, η οντότητα «cook\_credentials» χρησιμοποιείται για να συμβολίσει έναν πίνακα όπου θα αποθηκεύονται το username και το password κάθε μάγειρα και θα συνδέονται με το cook\_id του. Ωστόσο προκειμένου να επιδείξουμε την εργασία έχουμε αποθηκεύσει μία μόνο γραμμή στον πίνακα αυτό, η οποία αντιστοιχεί στις πληροφορίες ενός μόνο μάγειρα.

Αξίζει να αναφερθεί ότι, όπως φαίνεται και στο διάγραμμα και, σύμφωνα με τη μέθοδο aggregation , έχουμε θεωρήσει μία οντότητα η οποία περιλαμβάνει τις οντότητες «episodes\_per\_year» και «cooks», μαζί με τη σχέση «features\_contestant» που τις συνδέει.

## **Σχεσιακό Διάγραμμα**

Το σχεσιακό διάγραμμα που προκύπτει από τη σχεδίαση της βάσης μας είναι το παρακάτω:

Το αρχείο της εικόνας καθώς και το model που φτιάξαμε στο MySQL Workbench βρίσκεται στο repository της εργασίας, με όνομα *Relational\_Image.png* και *Relational.mwb*.

Στο διάγραμμα φαίνονται και τα ευρετήρια που χρησιμοποιεί κάθε πίνακας, καθώς και τα triggers που φτιάξαμε για καθέναν από αυτούς, για τα οποία θα μιλήσουμε στο μέρος 3.

# **Υλοποίηση της Βάσης**

## **2.1 DDL**

Η δημιουργία της βάσης γίνεται στο MySQL DDL script με όνομα *Create\_Cooking\_DB* (βρίσκεται στο repository). Σε αυτό το script δημιουργούνται όλα τα tables, τα triggers και τα indexes που υλοποιούν τη βάση. Θα περιγράψουμε πιο αναλυτικά την λειτουργία των triggers και των indexes στο μέρος 3.

Σε όλα τα tables έχουν οριστεί τα κατάλληλα constraints, ώστε να εξασφαλίζεται η ορθότητα της βάσης δεδομένων. Τα constraints αυτά είναι τα εξής: Primary Keys, Foreign Keys, χρήση CHECK για έλεγχο ακεραιότητας του πεδίου τιμών, περιορισμός NOT NULL.

Όπως αναφέραμε και προηγουμένως, θεωρούμε ότι ο κάθε μάγειρας μπορεί να εκτελέσει αποκλειστικά όλες τις συνταγές που ανήκουν σε εθνική κουζίνα στην οποία αυτός εξειδικεύεται. Για τον λόγο αυτό δεν ορίζουμε στο DDL script κάποιον πίνακα που να συνδέει τους μάγειρες με τις συνταγές που μπορούν να τους ανατεθούν.

## **2.2 DML**

Η διαχείριση (manipulation) των δεδομένων έγινε με τη χρήση δύο διαφορετικών

Triggers

Η χρησιμότητα των triggers που δημιουργήθηκαν περιγράφεται παρακάτω:

* **calculates\_age\_on\_insert**

Υπολογίζει την ηλικία ενός μάγειρα που εισάγεται στη βάση δυναμικά με βάση την ημερομηνία γέννησής του .

* **if\_age\_needs\_to\_be\_changed**

Υπολογίζει την ηλικία του μάγειρα σε περίπτωση που γίνει κάποιο update σε κάποιον μάγειρα, έτσι ώστε αν είχε γίνει λάθος στην ημερομηνία γέννησης του να μπορεί να διορθώνεται δυναμικά και η ηλικία του.

* **prevents\_more\_than\_3\_tips**

Εξασφαλίζει ότι δεν πρόκειται να εισαχθούν περισσότερα από 3 tips, δηλαδή χρηστικές συμβουλές, για την ίδια συνταγή. Σε περίπτωση που αυτό συμβεί θα προκαλέσει error.

* **calculate\_calories\_for\_recipe\_on\_insert\_recipe**

Υπολογίζει τις θερμίδες ανά μερίδα για τη συνταγή που εισάγεται στη βάση, με βάση τις θερμίδες και τις αναλογίες των υλικών που χρειάζεται και που έχουν ήδη εισαχθεί.

* **calculate\_calories\_for\_recipe\_on\_insert\_needs\_ingredient**

Ανανεώνει τις θερμίδες ανά μερίδα για μία συνταγή εάν προστεθεί κάποιο νέο υλικό που χρειάζεται η συνταγή.

* **common\_national\_cuisine**

Εξασφαλίζει ότι δεν πρόκειται να εισαχθεί σε κάποιο επεισόδιο συνδυασμός μάγειρα και συνταγής που δεν έχουν κοινή εθνική κουζίνα. Εάν συμβεί αυτό, τότε θα προκληθεί error.

* **duplicate\_national\_cuisine**

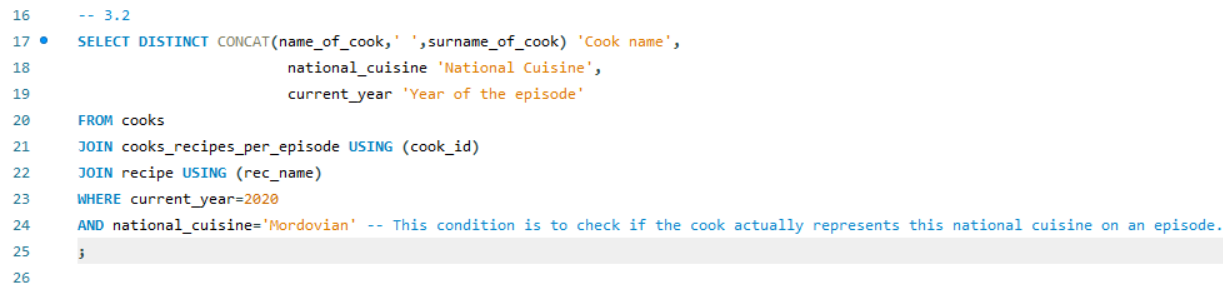
Εξασφαλίζει ότι δεν πρόκειται να εισαχθούν σε κάποιο επεισόδιο δύο συνταγές που ανήκουν στην ίδια εθνική κουζίνα. Εάν αυτό συμβεί θα προκληθεί error.

* **more\_than\_10\_cooks\_recipes**

Εξασφαλίζει ότι δεν πρόκειται να εισαχθούν στο ίδιο επεισόδιο πάνω από 10 συνδυασμοί συνταγών και μαγείρων. Θα προκαλέσει error στην περίπτωση που αυτό συμβεί.

Indexes/Ευρετήρια

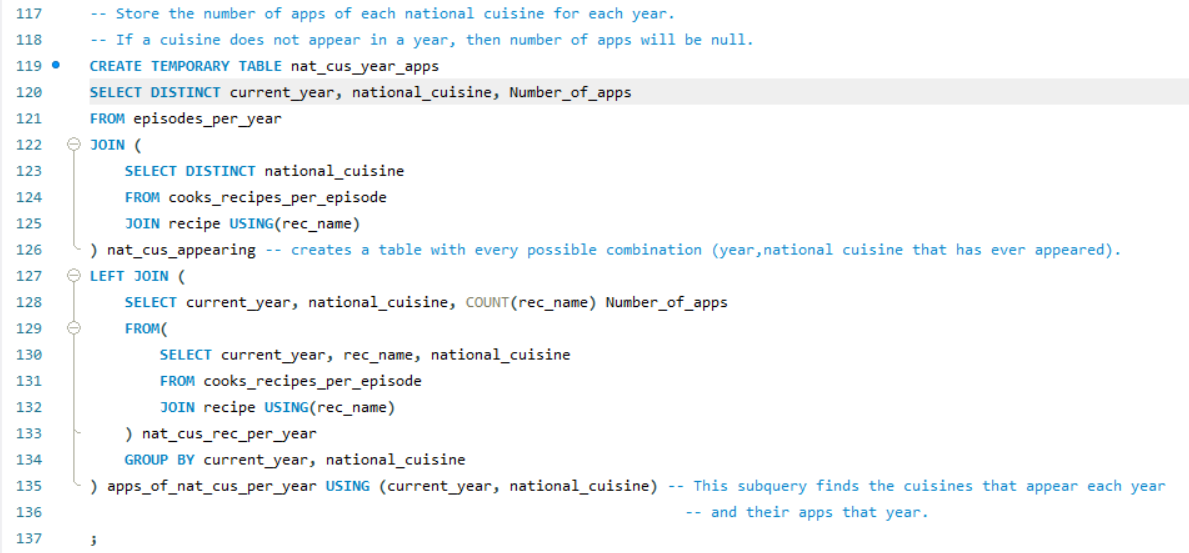
Όπως γνωρίζουμε, για όποια attributes έχουν καθοριστεί ως κλειδιά, είτε Primary Keys είτε Foreign Keys, δημιουργούνται αυτόματα και indexes από τη MySQL. Όπως είναι κατανοητό, τα περισσότερα queries θα βασίζονται σε αυτά τα attributes και συνεπώς δε θα απαιτούν επιπρόσθετα indexes. Υπάρχουν όμως κάποια queries που θα μπορούσαν να επωφεληθούν από τη χρήση κάποιων indexes. Συγκεκριμένα στο query 3.2 έχουμε τη συνθήκη: national\_cuisine = 'Ζητούμενη εθνική κουζίνα'.



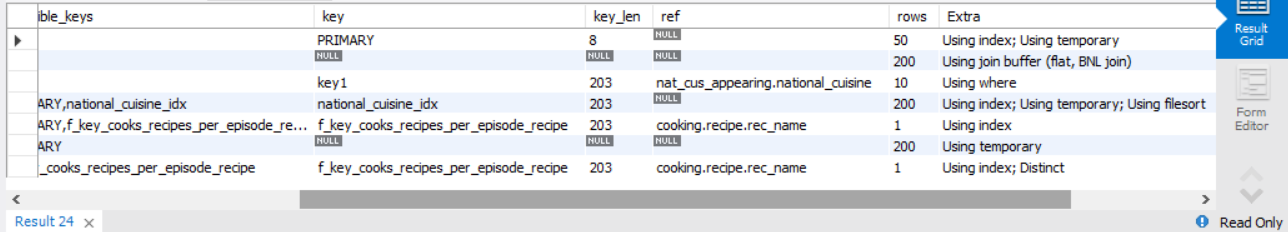
Τα indexes είναι κατάλληλα για τέτοιες συνθήκες αναζήτησης και συνεπώς θα ήταν καλή πρακτική να ορίσουμε ένα index για το table recipe στο attribute national\_cuisine.



Το συγκεκριμένο index χρησιμοποιείται και στο query 3.10, στο πρώτο subquery, ως συνθήκη JOIN.



Όπως μπορούμε να δούμε με τη βοήθεια της εντολής EXPLAIN, ο optimizer της SQL επιλέγει να χρησιμοποιήσει το συγκεκριμένο index.



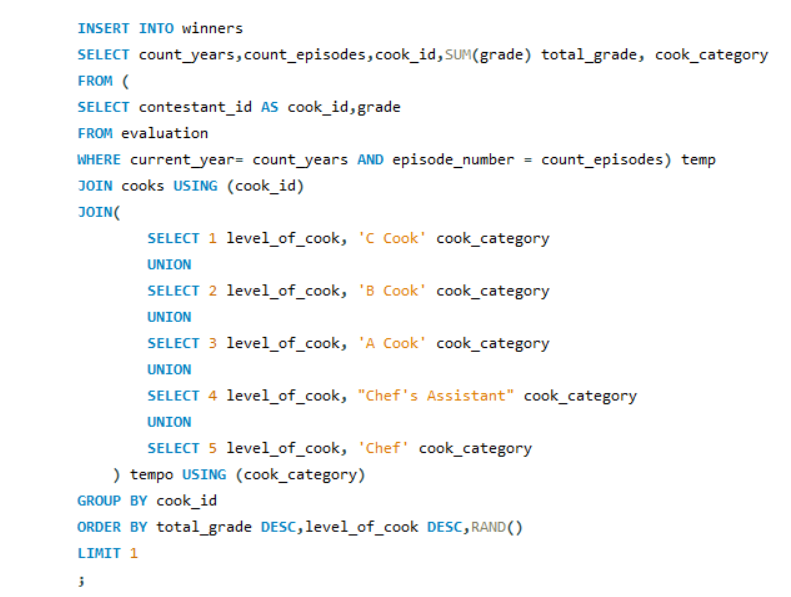
Επιπλέον, στο query 3.13 κάνουμε JOIN χρησιμοποιώντας το attribute cook\_category του cook.



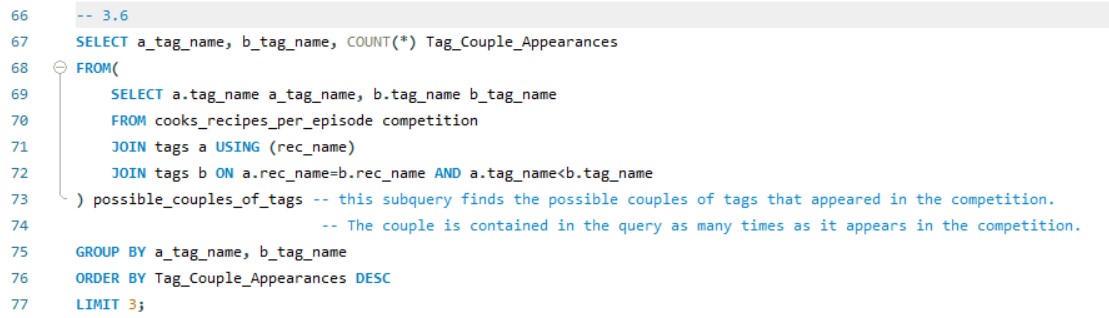
Μπορούμε να επισπεύσουμε τη διαδικασία του JOIN αυτού μέσω ενός index στο cook\_category:



Το συγκεκριμένο index, με την ίδια ακριβώς λογική, θα βοηθήσει και σε ένα subquery της κλήρωσης το οποίο επιλέγει τον νικητή του κάθε επεισοδίου:



Εναλλακτικά Query Plans

3.6) 

Σε αυτό το query ο optimizer της SQL επιλέγει να χρησιμοποιήσει το index του attribute rec\_name, τόσο στον πίνακα competition, όσο και στους δύο πίνακες tags. Η σειρά του JOIN επιλέγεται έτσι ώστε πρώτα να σκαναριστεί το table με τις λιγότερες γραμμές. Στα επόμενα tables του JOIN θα γίνει χρήση indexes και θα αποφευχθεί το σκανάρισμα των περισσότερων γραμμών. Παρακάτω φαίνονται τα αντίστοιχα traces που μπορούμε να δούμε μέσω της εντολής EXPLAIN και επιλέγοντας το FORMAT «JSON». Με κόκκινο φαίνονται τα indexes που χρησιμοποιούνται και με μπλε οι γραμμές που σκαναρίστηκαν από κάθε table. Όπως βλέπουμε το competition σκανάρεται 500 φορές (όσες και οι γραμμές του), ενώ τα δύο tables tags μόλις 2 φορές.

{

"query\_block": {

"select\_id": 1,

"filesort": {

"sort\_key": "count(0) desc",

"temporary\_table": {

"table": {

"table\_name": "competition",

"access\_type": "range",

"possible\_keys": ["f\_key\_cooks\_recipes\_per\_episode\_recipe"],

"key": "f\_key\_cooks\_recipes\_per\_episode\_recipe",

"key\_length": "203",

"used\_key\_parts": ["rec\_name"],

"rows": 500,

"filtered": 100,

"attached\_condition": "competition.rec\_name is not null and competition.rec\_name is not null",

"using\_index": true

},

"table": {

"table\_name": "a",

"access\_type": "ref",

"possible\_keys": ["PRIMARY", "f\_key\_tags\_recipe"],

"key": "f\_key\_tags\_recipe",

"key\_length": "202",

"used\_key\_parts": ["rec\_name"],

"ref": ["cooking.competition.rec\_name"],

"rows": 2,

"filtered": 100,

"using\_index": true

},

"table": {

"table\_name": "b",

"access\_type": "ref",

"possible\_keys": ["PRIMARY", "f\_key\_tags\_recipe"],

"key": "f\_key\_tags\_recipe",

"key\_length": "202",

"used\_key\_parts": ["rec\_name"],

"ref": ["cooking.competition.rec\_name"],

"rows": 2,

"filtered": 100,

"attached\_condition": "a.tag\_name < b.tag\_name",

"using\_index": true

}

}

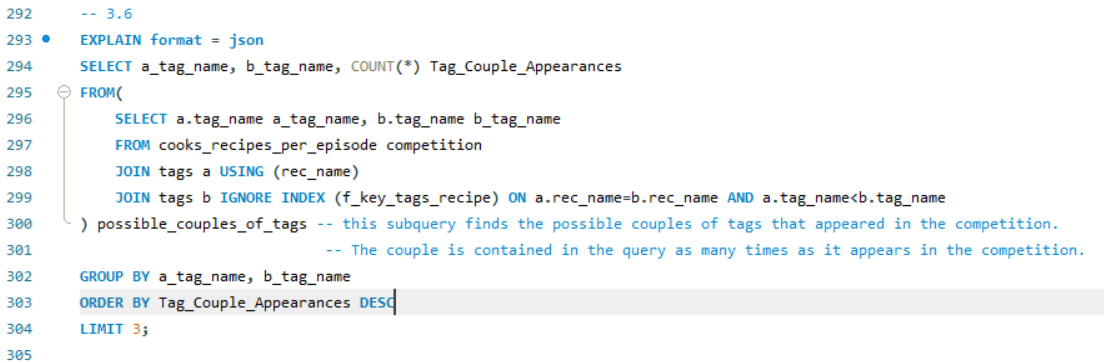
}

}

}

,

Εάν εξαναγκάσουμε τον optimizer της SQL να αγνοήσει το index του rec\_name, μέσω της εντολής IGNORE INDEX, στον πίνακα tags b, ο optimizer θα αναγκαστεί να χρησιμοποιήσει το index του Primary Key του tags, που όμως δεν είναι κατάλληλο στη συγκεκριμένη περίπτωση λόγω της συνθήκης: a.tag\_name < b.tag\_name.

 Για αυτό θα επιλέξει ως πρώτο table στο JOIN το tags b, παρόλο που έχει περισσότερες γραμμές από το competition. Έτσι καταφέρνει να εξετάσει αυτήν τη συνθήκη μόνο 2 φορές , για τις 2 γραμμές που ικανοποίησαν τα JOINS.

{

"query\_block": {

"select\_id": 1,

"filesort": {

"sort\_key": "count(0) desc",

"temporary\_table": {

"table": {

"table\_name": "b",

"access\_type": "index",

"possible\_keys": ["PRIMARY"],

"key": "PRIMARY",

"key\_length": "404",

"used\_key\_parts": ["tag\_name", "rec\_name"],

"rows": 1028,

"filtered": 100,

"using\_index": true

},

"table": {

"table\_name": "competition",

"access\_type": "ref",

"possible\_keys": ["f\_key\_cooks\_recipes\_per\_episode\_recipe"],

"key": "f\_key\_cooks\_recipes\_per\_episode\_recipe",

"key\_length": "203",

"used\_key\_parts": ["rec\_name"],

"ref": ["cooking.b.rec\_name"],

"rows": 1,

"filtered": 100,

"using\_index": true

},

"table": {

"table\_name": "a",

"access\_type": "ref",

"possible\_keys": ["PRIMARY", "f\_key\_tags\_recipe"],

"key": "f\_key\_tags\_recipe",

"key\_length": "202",

"used\_key\_parts": ["rec\_name"],

"ref": ["cooking.b.rec\_name"],

"rows": 2,

"filtered": 100,

"attached\_condition": "a.tag\_name < b.tag\_name",

"using\_index": true

}

}

}

}

}

Τέλος, ένα ακόμα ενδιαφέρον query plan θα ήταν να εξαναγκάσουμε επιπλέον τον optimizer να εκτελέσει τα JOINS με τη συγκεκριμένη σειρά που έχουν γραφτεί τα tables στον κώδικα. Αυτό το επιτυγχάνουμε μέσω της εντολής STRAIGHT\_JOIN.

photo

Αυτή τη φορά ο optimizer δεν μπορεί να αποφύγει τη χρήση του index (tag\_name,rec\_name) του Primary Key για το JOIN με τη συνθήκη a.tag\_name < b.tag\_name. Έτσι καταφεύγει στη χρήση της τεχνικής Block Nested Loop Join. Δηλαδή αποθηκεύει σε κάθε εξωτερική λούπα σε έναν buffer μερικά υποψήφια tuples των άλλων 2 tables και τα εξετάζει ένα προς ένα στην εσωτερική λούπα. Τα traces φαίνονται παρακάτω:

{

"query\_block": {

"select\_id": 1,

"filesort": {

"sort\_key": "count(0) desc",

"temporary\_table": {

"table": {

"table\_name": "competition",

"access\_type": "range",

"possible\_keys": ["f\_key\_cooks\_recipes\_per\_episode\_recipe"],

"key": "f\_key\_cooks\_recipes\_per\_episode\_recipe",

"key\_length": "203",

"used\_key\_parts": ["rec\_name"],

"rows": 500,

"filtered": 100,

"attached\_condition": "competition.rec\_name is not null",

"using\_index": true

},

"table": {

"table\_name": "a",

"access\_type": "ref",

"possible\_keys": ["PRIMARY", "f\_key\_tags\_recipe"],

"key": "f\_key\_tags\_recipe",

"key\_length": "202",

"used\_key\_parts": ["rec\_name"],

"ref": ["cooking.competition.rec\_name"],

"rows": 2,

"filtered": 100,

"using\_index": true

},

"block-nl-join": {

"table": {

"table\_name": "b",

"access\_type": "index",

"possible\_keys": ["PRIMARY"],

"key": "PRIMARY",

"key\_length": "404",

"used\_key\_parts": ["tag\_name", "rec\_name"],

"rows": 1028,

"filtered": 100,

"using\_index": true

},

"buffer\_type": "flat",

"buffer\_size": "256Kb",

"join\_type": "BNL",

"attached\_condition": "b.rec\_name = competition.rec\_name and a.tag\_name < b.tag\_name"

}

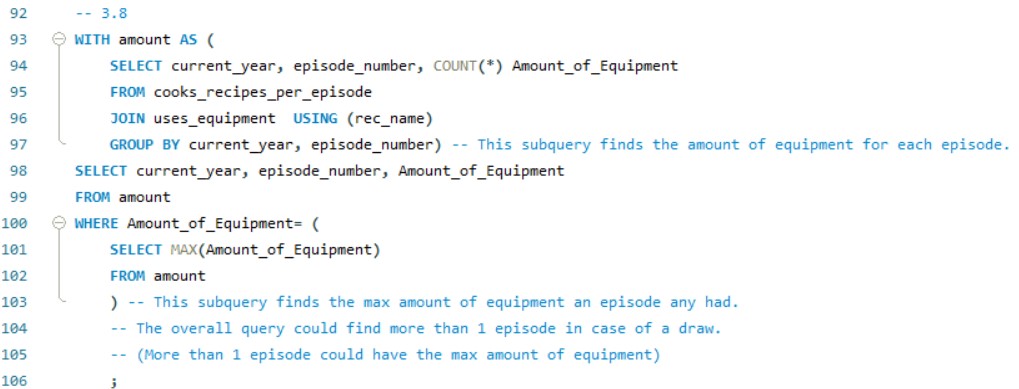
}

}

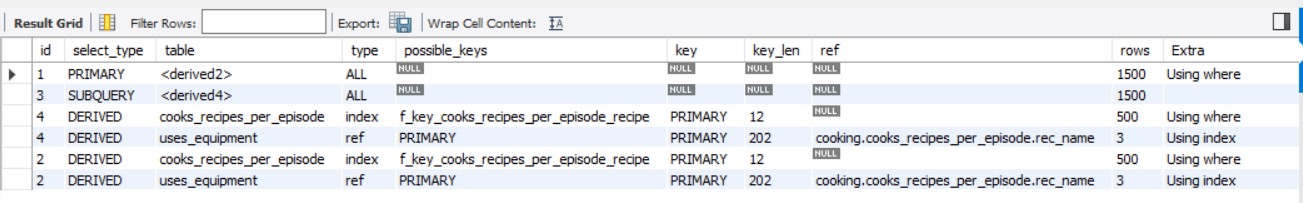
}

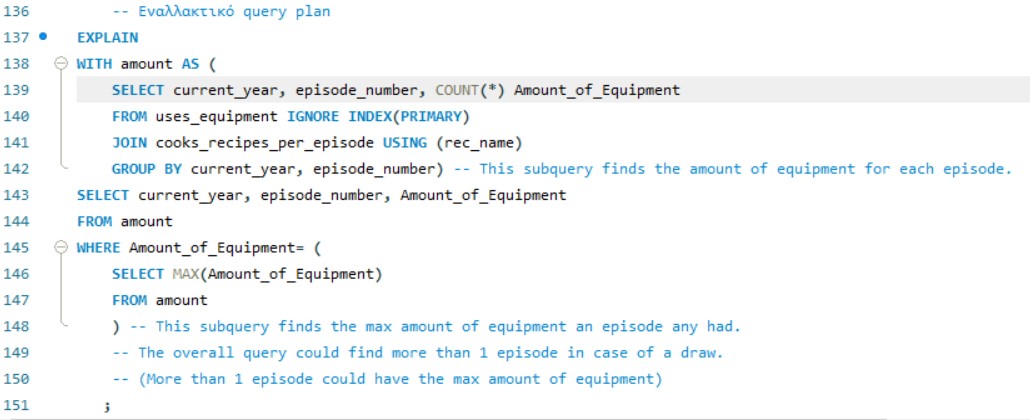
}

Φυσικά η συγκεκριμένη παρέμβαση μειώνει αρκετά την αποδοτικότητα του query, όπως φαίνεται από τις 1028 γραμμές που σκανάρονται στο τελευταίο table.

3.8) 

Σε αυτό το query ο optimizer επιλέγει να κάνει το JOIN του πρώτου subquery χρησιμοποιώντας το Primary Key του uses\_equipment, δηλαδή index στο rec\_name. Έτσι σκανάρει πρώτα το table cooks\_recipes\_per\_episode που έχει λιγότερες γραμμές (500 όπως φαίνεται παρακάτω) και στη συνέχεια χρησιμοποιεί το index για το uses\_equipment, σκανάροντας μόνο 3 γραμμές.



Ένα εναλλακτικό query plan θα ήταν να επιβάλλουμε στον optimizer να αγνοήσει το Primary Key του uses\_equipment. 

Αυτό θα οδηγήσει τον optimizer στο να χρησιμοποιήσει το index στο rec\_name του cooks\_recipes\_per\_episode, σκανάροντας πρώτα το uses\_equipment, το οποίο όπως φαίνεται παρακάτω έχει 1393 γραμμές, μειώνοντας την αποδοτικότητα του query.

