ΠΑΝΕΠΙΣΤΗΜΙΟ ΙΩΑΝΝΙΝΩΝ

ΤΜΗΜΑ ΜΗΧ. Η/Υ & ΠΛΗΡΟΦΟΡΙΚΗΣ

Προχωρημένα Θέματα Τεχνολογίας και Εφαρμογών Βάσεων Δεδομένων

ΠΡΟΓΡΑΜΜΑΤΙΣΤΙΚΗ ΕΡΓΑΣΙΑ για το ακαδημαϊκο έτος 2021-2022

Χαρισησ αλεξανδροσ, 3361

δημητρησ βαμπιρησ 3186

κωνσταντινοσ ζιουδασ 3225

ΤΕΛΙΚΗ ΑΝΑΦΟΡΑ

ΜΑΪΟΣ 2022

# βαση δεδομενων

Στην παρούσα άσκηση τα δεδομένα δίνονται με ένα συγκεκριμένο φορμάτ:

Table

Description automatically generated

Εικόνα 1: Μικρό απόσπασμα από τα στατιστικά της Αργεντίνας.

Παρατηρούμε πως χρειάζεται κανονικοποίηση των δεδομένων για μεγαλύτερη ευκολία στη χρήση και για την αποφυγή σφαλμάτων.

Σκοπός του τι θέλουμε να υλοποιήσουμε είναι η προβολή ή η σύγκριση, πολλών indicator ή χωρών μεταξύ τους, οπότε αποφασίσαμε να μετατρέψουμε και να αποθηκεύσουμε τα δεδομένα στο τραπέζι ‘stats’ στην παρακάτω μορφή:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Country\_id | Year | Indicator1 | Indicator2 | … | IndicatorN |
| Country\_id0 | minYear | value | value | … | value |
| Country\_id0 | minYear+1 | value | value | … | value |
| Country\_id0 | … | value | value | … | value |
| Country\_id0 | maxYear | value | value | … | value |
| Country\_id1 | minYear | value | value | … | value |
| Country\_id1 | minYear+1 | value | value | … | value |
| Country\_id1 | … | value | value | … | value |
| Country\_id1 | maxYear | value | value | … | value |
| …. | …. | … | … | … | … |
| Country\_idN | minYear | value | value | … | value |
| Country\_idN | minYear+1 | value | value | … | value |
| Country\_idN | … | value | value | … | value |
| Country\_idN | maxYear | value | value | … | value |

Εικόνα 2: Stats table format

Κύριο κλειδί αποτελεί η χώρα και η χρονιά, συμβάλοντας στη κάθε χώρα να εμφανίζεται N φορές, όσα και τα χρόνια. Κάθε συνδυασμός χώρας και χρόνου συνοδεύεται από τις τιμές του κάθε indicator (index) για εκείνη τη συγκεκριμένη χρονιά και χώρα.

Προσφέρουμε κατ’ αυτόν τον τρόπο την δυνατότητα εύρεσης πολλαπλών (ή όλων) των στατιστικών που συσχετίζονται με μία χρονιά και μία χώρα, γεγονός που μας επιτρέπει πιο εύκολα να συγκρίνουμε πολλά indicator την ίδια στιγμή.

Ανάλυση του τραπεζιού:

Κάθε κελί αποτελείται από μοναδική τιμή με μοναδικούς συνδυασμούς κλειδιών, οπότε το τραπέζι είναι μορφής 1NF.

Έπειτα, επειδή και η χρονιά αλλά και η χώρα καθορίζουν την τιμή μίας ένδειξης, το τραπέζι βρίσκεται και σε μορφή 2NF.

Τέλος, καμία στήλη από τα indicators δεν εξαρτάται από άλλη πέρα του primary key.

Άρα το τραπέζι μας stats είναι 3NF. Αυτό σημαίνει πως είναι απαλλαγμένο από ανωμαλίες εισαγωγής, ενημέρωσης και διαγραφής των δεδομένων της.

Βοηθητικά τραπέζια:

Για να είναι πιο κατηγοριοποιημένη η σχεδίαση αλλά και να έχουμε στη διάθεσή μας περισσότερες πληροφορίες, δημιουργήσαμε δύο βοηθητικά τραπέζια.

Πρώτο και πιο σχετικό τραπέζι είναι το ‘countries’, το οποίο περιέχει κάθε χώρα ως όνομα, κωδικό, country\_id και λοιπές πληροφορίες ή σχόλια.

Table

Description automatically generated

Το country\_id αποτελεί το primary key, και αυτό περνιέται και ως foreign key στο τραπέζι stats. Αποφασίσαμε να δηλώσουμε αυτό ως κλειδί διότι αποτελεί έναν απλό αριθμό και στο φόρτωμα όλου του τραπεζιού ‘stats’ χρειάζεται διασταύρωση του εκείνου του κλειδιού με του τραπεζιού ‘countries’, πράγμα που το καθιστά πιο γρήγορο άμα είναι αριθμητικό.

Επίσης είναι λιγότερο πιθανό να έχει τον ίδιο κωδικά με άλλη χώρα, όπως κάνει το country\_code που αποτελείται από 3 γράμματα και έχει περιορισμένους συνδυασμούς.

Στη συνέχεια θα δούμε το μεμονομένο τραπέζι ‘indicators’ το οποίο συγκρατεί πληροφορίες σχετικά με τα indicators όπως όνομα, κωδικό κλπ. Κι εκεί έχουμε βάλει ως primary key το indicator\_id για να αποφύγουμε τη χρήση μεγάλων αλφαρηθμιτικών.

Graphical user interface, text, application

Description automatically generated

## Σχεσιακό σχήμα σε λογικό επίπεδο

Graphical user interface, application, table, Excel

Description automatically generated

Σχήμα 1.1 Σχεσιακό σχήμα της βάσης δεδομένων του συστήματος

Database creation command:

CREATE DATABASE `dbms` /\*!40100 DEFAULT CHARACTER SET utf8mb4 COLLATE utf8mb4\_0900\_ai\_ci \*/ /\*!80016 DEFAULT ENCRYPTION='N' \*/;

Countries table creation command:

CREATE TABLE `countries` (

`country\_code` varchar(3) DEFAULT NULL,

`region` varchar(64) DEFAULT NULL,

`income\_group` varchar(64) DEFAULT NULL,

`country\_name` varchar(64) DEFAULT NULL,

`special\_notes` text,

`country\_id` int NOT NULL AUTO\_INCREMENT,

PRIMARY KEY (`country\_id`)

) ENGINE=InnoDB AUTO\_INCREMENT=26 DEFAULT CHARSET=utf8mb4 COLLATE=utf8mb4\_0900\_ai\_ci;

Indicators table creation command:

CREATE TABLE `indicators` (

`indicator\_code` varchar(32) DEFAULT NULL,

`indicator\_name` text,

`source\_note` text,

`source\_organization` text,

`indicator\_id` int NOT NULL AUTO\_INCREMENT,

PRIMARY KEY (`indicator\_id`)

) ENGINE=InnoDB AUTO\_INCREMENT=56 DEFAULT CHARSET=utf8mb4 COLLATE=utf8mb4\_0900\_ai\_ci;

Stats table creation command:

CREATE TABLE `stats` (

`country\_id` int NOT NULL,

`year` year NOT NULL,

`gc\_tax\_expt\_cn` double DEFAULT NULL,

`eg\_fec\_rnew\_zs` double DEFAULT NULL,

`bm\_gsr\_mrch\_cd` double DEFAULT NULL,

`gc\_tax\_ypkg\_cn` double DEFAULT NULL,

`eg\_elc\_rnew\_zs` double DEFAULT NULL,

`bm\_gsr\_fcty\_cd` double DEFAULT NULL,

`gc\_tax\_othr\_rv\_zs` double DEFAULT NULL,

`eg\_elc\_ngas\_zs` double DEFAULT NULL,

`gc\_tax\_gsrv\_va\_zs` double DEFAULT NULL,

`eg\_use\_comm\_cl\_zs` double DEFAULT NULL,

`eg\_elc\_accs\_ur\_zs` double DEFAULT NULL,

`bm\_gsr\_tran\_zs` double DEFAULT NULL,

`gc\_tax\_impt\_cn` double DEFAULT NULL,

`eg\_use\_comm\_fo\_zs` double DEFAULT NULL,

`eg\_elc\_accs\_zs` double DEFAULT NULL,

`bm\_gsr\_trvl\_zs` double DEFAULT NULL,

`gc\_tax\_totl\_gd\_zs` double DEFAULT NULL,

`eg\_elc\_petr\_zs` double DEFAULT NULL,

`bm\_gsr\_cmcp\_zs` double DEFAULT NULL,

`gc\_tax\_othr\_cn` double DEFAULT NULL,

`eg\_use\_pcap\_kg\_oe` double DEFAULT NULL,

`eg\_elc\_loss\_zs` double DEFAULT NULL,

`bm\_trf\_pwkr\_cd\_dt` double DEFAULT NULL,

`gc\_tax\_ypkg\_rv\_zs` double DEFAULT NULL,

`eg\_elc\_rnwx\_kh` double DEFAULT NULL,

`bm\_gsr\_gnfs\_cd` double DEFAULT NULL,

`gc\_tax\_intt\_rv\_zs` double DEFAULT NULL,

`eg\_use\_elec\_kh\_pc` double DEFAULT NULL,

`eg\_elc\_hyro\_zs` double DEFAULT NULL,

`bm\_trf\_prvt\_cd` double DEFAULT NULL,

`gc\_tax\_intt\_cn` double DEFAULT NULL,

`eg\_use\_crnw\_zs` double DEFAULT NULL,

`eg\_elc\_fosl\_zs` double DEFAULT NULL,

`bm\_klt\_dinv\_wd\_gd\_zs` double DEFAULT NULL,

`gc\_tax\_gsrv\_rv\_zs` double DEFAULT NULL,

`eg\_imp\_cons\_zs` double DEFAULT NULL,

`eg\_elc\_accs\_ru\_zs` double DEFAULT NULL,

`bm\_gsr\_totl\_cd` double DEFAULT NULL,

`gc\_tax\_expt\_zs` double DEFAULT NULL,

`eg\_gdp\_puse\_ko\_pp` double DEFAULT NULL,

`eg\_cft\_accs\_zs` double DEFAULT NULL,

`bm\_gsr\_nfsv\_cd` double DEFAULT NULL,

`gc\_tax\_gsrv\_cn` double DEFAULT NULL,

`eg\_gdp\_puse\_ko\_pp\_kd` double DEFAULT NULL,

`eg\_egy\_prim\_pp\_kd` double DEFAULT NULL,

`bm\_gsr\_royl\_cd` double DEFAULT NULL,

`gc\_tax\_totl\_cn` double DEFAULT NULL,

`eg\_elc\_nucl\_zs` double DEFAULT NULL,

`gc\_tax\_ypkg\_zs` double DEFAULT NULL,

`eg\_elc\_rnwx\_zs` double DEFAULT NULL,

`bm\_gsr\_insf\_zs` double DEFAULT NULL,

`gc\_tax\_impt\_zs` double DEFAULT NULL,

`eg\_use\_comm\_gd\_pp\_kd` double DEFAULT NULL,

`eg\_elc\_coal\_zs` double DEFAULT NULL,

`bm\_klt\_dinv\_cd\_wd` double DEFAULT NULL,

PRIMARY KEY (`country\_id`,`year`),

CONSTRAINT `stats\_ibfk\_1` FOREIGN KEY (`country\_id`) REFERENCES `countries` (`country\_id`)

) ENGINE=InnoDB DEFAULT CHARSET=utf8mb4 COLLATE=utf8mb4\_0900\_ai\_ci;

## Σχεσιακό σχήμα σε φυσικο επίπεδο

### Ρύθμιση των παραμέτρων του dbms

storage engine,memory allocation (of various kinds), …

Η InnoDB είναι μια μηχανή αποθήκευσης που εξισορροπεί την υψηλή αξιοπιστία και την υψηλή απόδοση. Συγκεκριμένα έχει μεγαλύτερη ταχύτητα γραφής, γεγονός που συμφέρει δεόντος στην περίπτωσή μας που φορτώνουμε πολλά δεδομένα από εξωτερικά αρχεία.

Γαι την επίτρεψη φόρτωσης από τοπικά αρχεία υπήρχαν διάφορες αλλαγές που έπρεπε να γίνουν:

1. Εκτέλεση της mysql εντολής: SET GLOBAL local\_infile = ‘ON’;  
   (Αυτό γίνεται μέσω του κώδικα).
2. Τερματισμός της mysql μέσω του terminal με την εντολή: net stop MySQL80
3. Άνοιγμα της επιλογής εμφάνισης hidden files,
4. Άνοιγμα του αρχείου‘’ C:\ProgramData\MySQL\MySQL Server 8.0\my.ini‘’ και
5. Αλλαγή της μεταβλητής secure-file-priv σε “” (0-length string).
6. Έναρξη της mysql μέσω του terminal με την εντολή: net start MySQL80

Τέλος, θέτουμε το μέγεθος του buffer της mysql στο ιδανικότερο για την εφαρμογή, εκτελώντας την mysql εντολή:

* SET GLOBAL innodb\_buffer\_chuck\_size = 1.048.576;
* SET GLOBAL innodb\_buffer\_pool\_size = 1.048.576;

### Ρύθμιση του φυσικού σχήματος της βάσης δεδομένων

Όπως υπώθηκε [προηγουμένως](#σελίδα2) έχουμε δημιουργήσει τη βάση με τον σκοπό της στο μυαλό μας, αντιπροσωπεύoντάς την καλύτερα στις ανάγκες/απαιτήσεις της εφαρμογής.

### Ρύθμιση ασφάλειας

Έχει δημιουργηθεί ένας backup φάκελος με τα τραπέζια του dmbs ως .csv αρχεία, μέσω του export wizard του mysql workspace.

# αρχιτεκτονικη Λογισμικού

## αρχιτεκτονικη και δομη ETL

Diagram

Description automatically generated with medium confidence

Text

Description automatically generatedΕικόνα 3: Αρχιτεκτονική φόρτωσης και μετατροπής δεδομένων

Κάθε .zip αρχείο που κατεβάζαμε περιείχε 3 csv αρχεία, το καθένα με διαφορετικά δεδομένα. Τα χωρίσαμε σε 3 φακέλους βάση κατηρογίας.

Εικόνα : /dbms/data/original/

Έτσι κάθε χώρα έχει από ένα αρχείο σε κάθε αντιπροσωπευτικό φάκελο.

Βάση κάθε φακέλου παράγουμε ένα αντίστοιχο τελικό αρχείο, περιέχοντας όλες τις ζητούμενες πληροφορίες από τα αρχεία εκείνου του φακέλου.

Η διαδικασία μετατροπής περιγράφεται παραπάνω.

Χειροκίνητες αλλαγές στο σύστημα:

Για τη μετατροπή των αρχείων σε csv τελικής μορφής δεν απαιτούνται αλλαγές στον κώδικα.

Ωστόσο για τη σύνδεση στην βάση χρειάζεται αλλαγή των παραμέτρω σύνδεσης, δεδομένου ότι ο χρήστης χρησιμοποιεί διαφορετικό user & password.

Αρκεί ο χρήστης να πάει στον φάκελο ‘../DBMS/src/’ στο αρχείο settings.py και να βάλει στα user, password (και άμα θέλει database) τα δικά του δεδομένα.

## Διαγράμματα ΠΑΚΕΤΩΝ / υποσυστημάτων kentρικησ εφαρμογησ

Το διάγραμμα για τα υποσυστήματα / πακέτα του λογισμικού που κατασκευάσατε ως κεντρική εφαρμογή επερώτησης. Ο στόχος είναι να φανεί η high-level αρχιτεκτονική του συστήματος, χωρίς λεπτομέρειες των επί μέρους κλάσεων. Κάποιος πολύ σύντομος σχολιασμός επίσης.

Diagram

Description automatically generated

Εικόνα 5: High-level αρχιτεκτονική του συστήματος / packages.

## Διαγράμμα(τα) Κλάσεων κεντρικησ εφαρμογησ

Έχει χρησιμοποιηθεί αρχιτεκτονική Model-Controller-View για διαχωρισμό υπευθυνοτήτων και πιο εύκολη ανάπτυξη του συστήματος.

Diagram, text

Description automatically generated

Εικόνα : Main & settings

Diagram

Description automatically generated

Εικόνα 7: Dataload

Letter

Description automatically generated with medium confidence

Εικόνα 8: Client & Controller

Graphical user interface

Description automatically generated with medium confidence

Εικόνα 9: Plotter, databaseConnector & dataFormatter

Diagram

Description automatically generated

Εικόνα 10: Model package interactions

# Υποδείγματα ερωτήσεων και απαντήσεων

1. Κατεύθυνση στο directory …/DMBS/src/ και άνοιγμα της εφαρμογής με την εκκίνηση του main.py.
2. Ο χρήστης επιλέγει τις χώρες, χρονιές και στατιστικές της επιλογής του και του ανοίγουν οι αντίστοιχες επιλογές για δημιουργία plot.
3. Δυνατές επιλογές:
   1. Timeline Plots: k-στατιστικές και k’-χώρες
   2. Bar Plots: k-στατιστικές και k’-χώρες
   3. Scatter Plots: 2 στατιστικές και
      1. Ίδια χρονιά, καμία χώρα. (δείχνει για όλες μαζί εκείνη τη χρονιά)
      2. Διαφορετικές χρονιές, μία χώρα

Graphical user interface, application

Description automatically generated

Εικόνα 11: Αρχική κατάστασης της εφαρμογής

Graphical user interface, text, application

Description automatically generated

Εικόνα 12: Επιλογή μεταβλητών και ενεργοποίηση ανάλογων plot επιλογών.

Graphical user interface

Description automatically generated

Εικόνα 13: Επιλογή 2 χωρών και μίας χρονιάς δίχως χώρα για την ενεργοποίηση scatter

# λοιπά σχόλια

Τα Tests τρέχουν με το τρέξιμο του tests.py, το οποίο βρίσκεται στον ίδιο φάκελο με την main.