ΠΑΝΕΠΙΣΤΗΜΙΟ ΙΩΑΝΝΙΝΩΝ

ΤΜΗΜΑ ΜΗΧ. Η/Υ & ΠΛΗΡΟΦΟΡΙΚΗΣ

2023-2024

**Προχωρημένα Θέματα Τεχνολογίας**

**και**

**Εφαρμογών Βάσεων Δεδομένων**

ΑΝΑΣΤΑΣΙΑ ΒΑΡΟΥΧΑ – 4288

ΚΑΤΕΡΙΝΑ ΠΑΠΑΧΡΙΣΤΟΠΟΥΛΟΥ - 4148

ΜΑΪΟΣ 2024

**Project Progress**

|  |  |  |  |
| --- | --- | --- | --- |
| Ημερομηνία | Έκδοση | Περιγραφή | Συγγραφέας |
| 2024/02/24 | 0.1 | Set up the necessary tools for project. Navigate through csv files. | 4288, 4148 |
| 2024/03/20 | 0.1 | Make the first script that will transform the data. | 4288, 4148 |
| 2024/03/26 | 0.2 | Fix the script, and set up the db. | 4288, 4148 |
| 2024/03/30 | 0.2 | Load the new CVSs to db. Check the db. Ask questions and see the results. | 4288, 4148 |
| 2024/04/10 | 0.3 | Make the first chart scripts. | 4288, 4148 |
| 2024/04/15 | 0.3 | Set up the environment for our Flask app. Try to connect the db to it. | 4288, 4148 |
| 2024/04/22 | 0.3 | Refactor the scripts and add them to our app. | 4288, 4148 |
| 2024/04/30 | 0.4 | Make the first UI. | 4288, 4148 |
| 2024/05/05 | 0.4 | Successfully connect the db to the app. | 4288, 4148 |
| 2024/05/10 | 0.5 | Add the aggregation feature. | 4288, 4148 |
| 2024/05/15 | 0.6 | Add the filtering feature. | 4288, 4148 |
| 2024/05/20 | 0.7 | Make the UI better-looking and refactor it. | 4288, 4148 |
| 2024/05/22 | 0.8 | Add info button to guide the client. | 4288, 4148 |
| 2024/05/25 | 0.9 | Export data for certain countries. | 4288, 4148 |
| 2024/05/26 | 1.0 | Make the report , README. Upload the project to GitHub. | 4288, 4148 |
| 2024/05/28 | 1.0 | Record demo. | 4288, 4148 |
| 2024/05/30 | 1.0 | Make last changes and reupload the app. | 4288, 4148 |

ΜΑΪΟΣ 2024

**Περιγραφή Προβλήματος**

Η προγραμματιστική άσκηση επικεντρώνεται στην αντιμετώπιση του προβλήματος της **ενοποίησης** και **οπτικοποίησης** δεδομένων. Συγκεκριμένα, στοχεύει στην ενοποίηση δεδομένων με διαφορετικές δομές και κωδικοποιήσεις σε ένα ενιαίο σχήμα, καθιστώντας τα επερωτήσιμα από μια βάση δεδομένων. Επιπλέον, επιδιώκεται η ανάπτυξη τεχνικών οπτικοποίησης που θα αναδείξουν **οπτικά κρυμμένες ιδιότητες,** **τάσεις** και **πρότυπα** στα δεδομένα, παρέχοντας στους χρήστες πληροφορίες με τρόπο κατανοητό και δια δραστικό.

**Σκοπός**

**Στόχος** του project είναι η δημιουργία μιας εφαρμογής που θα **ενσωματώνει δεδομένα δημογραφικών** και **οικονομικών** στοιχείων σε μια ενιαία βάση δεδομένων και θα επιτρέπει τη δια δραστική πλοήγηση σε αυτά. Μέσω της εφαρμογής, ο χρήστης θα μπορεί να επιλέξει **χώρες, δείκτες** και **χρονικά διαστήματα** για να εξετάσει την εξέλιξη των δεικτών στο χρόνο, να συγκρίνει μετρήσεις και να ανακαλύψει συσχετίσεις μεταξύ διαφορετικών μεγεθών.

ΜΑΪΟΣ 2024

**1.1 Σχεσιακό σχήμα της Βάσης σε λογικό επίπεδο**

A screenshot of a computer

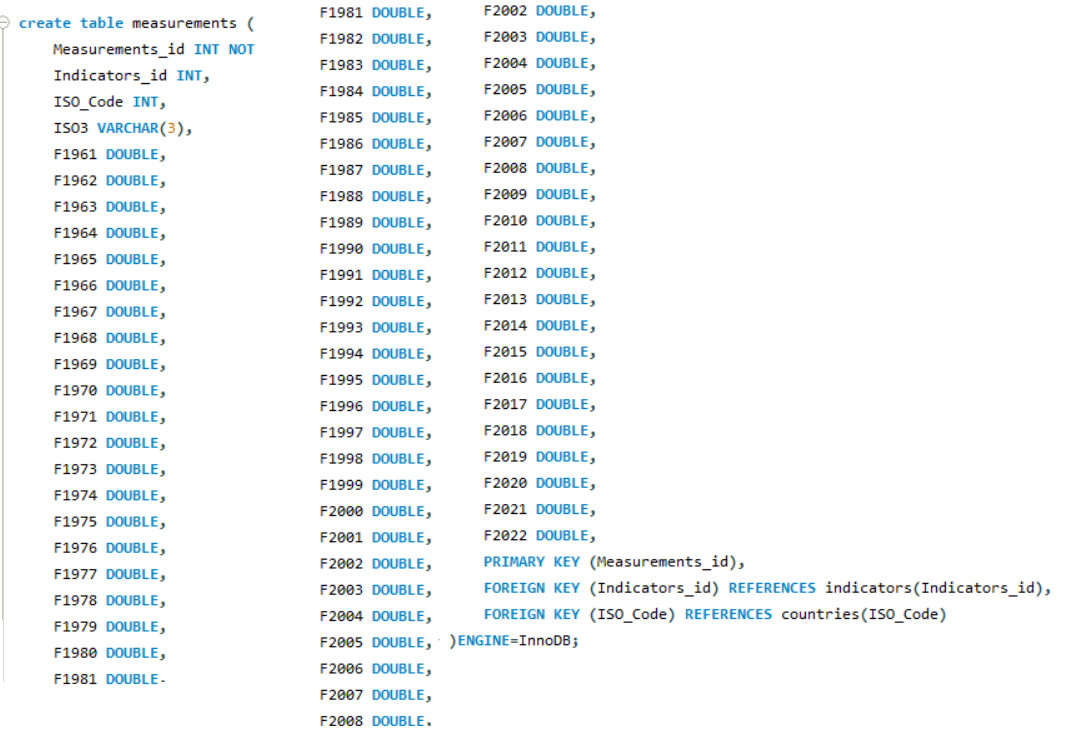
Description automatically generated

ΜΑΪΟΣ 2024

* 1. **Σχεσιακό σχήμα της Βάσης του συστήματος:**

Εικόνα που περιέχει κείμενο, στιγμιότυπο οθόνης, γραμματοσειρά

Περιγραφή που δημιουργήθηκε αυτόματα



ΜΑΪΟΣ 2024

* 1. **Σχεσιακό σχήμα σε φυσικό επίπεδο:**

**1.2.1. Ρυθμίσεις παραμέτρων του DBMS:**

Για να κάνουμε σωστό **set** του περιβάλλοντος της βάσης μας πρέπει να ξεκινήσουμε με ορισμένες τροποποιήσεις στο περιβάλλον της **MySQL**. Κάτι πολύ σημαντικό είναι να ελέγξουμε την συνολική κύρια μνήμη του μηχανήματος που θα φτιαχτεί σε πρώτο χρόνο η βάση και στην συνέχεια η εφαρμογή. H κύρια μνήμη που διαθέτει το συγκεκριμένο μηχάνημα συνολικά είναι **7.9GB** όπου τα **4.1GB** από αυτά δεσμεύονται από άλλες αναγκαίες διεργασίες. Συνεπώς εμείς θα θεωρήσουμε πως η ελεύθερη κύρια μνήμη που διαθέτει το **dedicated server** μηχάνημα είναι **3.8 GB** και πρέπει να αναθέσουμε στην **MySQL** περίπου **2.66GB-3.04GB**, δηλαδή το 70-80% της ελεύθερης κύριας μνήμης. Συγκεκριμένα θέλουμε το **value** να ανέρχεται στα **268435456.** Για να αναθέσουμε το εν λόγο ποσοστό στην **MySQL** πρέπει να κάνουμε μερικά σημαντικά βήματα. Ανοίγουμε το workbench και γράφουμε **SET GLOBAL innodb\_buffer\_pool\_size = 268435456**;

Έτσι ώστε να αλλάξουμε από εκεί δυναμικά το **buffer\_pool value**.

Στην συνέχεια, ανοίγουμε το workbench της **MySQL** και ήρθε η ώρα να επιβεβαιώσουμε τις αλλαγές που κάναμε. Μπορούμε μέσο του query **SHOW VARIABLES LIKE 'innodb\_buffer\_pool\_size';** να μάθουμετο μέγεθος του **buffer pool**.

**Εικόνα που περιέχει κείμενο, στιγμιότυπο οθόνης, λογισμικό, εικονίδιο υπολογιστή

Περιγραφή που δημιουργήθηκε αυτόματα**

ΜΑΪΟΣ 2024

Κάτι ακόμα που θα χρειαστεί να κάνουμε είναι να πάμε στο directory : **Database -> Manage Connections->Select Local Server -> Select System Profile Tab**. Με αυτό τον τρόπο βρίσκουμε το **configuration file** .Εφόσον το λειτουργικό μας σύστημα είναι τα **Windows**, το **configuration file** έχει την κατάληξη **.ini**.

**Εικόνα που περιέχει κείμενο, στιγμιότυπο οθόνης, λογισμικό, αριθμός

Περιγραφή που δημιουργήθηκε αυτόματα**

Επίσης μπορούμε να αλλάξουμε το **default-storage-engine = INNODB.**

ΜΑΪΟΣ 2024

**2.1 ELT PROCESS (Extract – Transform – Load)**

Αφού έχουμε φτιάξει την βάση μας, έχουμε κάνει τις προσαρμογές στο περιβάλλον και έχουμε δημιουργήσει τα **tables** ήρθε η ώρα να φορτώσουμε τα δεδομένα από τα **csv**. Ωστόσο τα δεδομένα είναι αρκετά τροποποιημένα σε σχέση με τα αρχικά δεδομένα που μας δόθηκαν. Η διαδικασία χρήζει απαραίτητο να δημιουργήσουμε ένα **script** τροποποίησης με όνομα **modify\_data** το οποίο θα μας βοηθήσει να φτιάξουμε τα δεδομένα μας πιο καθαρά αλλά και πιο λειτουργικά έτσι ώστε να χρησιμοποιηθούν στην μετέπειτα διαδικασία της εφαρμογής μας.

Πάμε να δούμε τον τρόπο σκέψης μας πίσω από το **script**:

Έχουμε 5 αρχεία που καλούμαστε να διαχειριστούμε εκ των οποίων τα 4 περιέχουν μετρήσεις για περιβαλλοντικά φαινόμενα για πολλές χρονιές, για διάφορες χώρες του κόσμου. Επίσης , έχουμε το **countries.csv** το οποίο περιέχει στοιχεία ταυτότητας για κάποιες από αυτές τις χώρες. Σκοπός μας είναι να δημιουργήσουμε 3 τελικά αρχεία. Τα **countries.csv** , **indicators.csv** και **measurements.csv** όπου το κάθε ένα θα περιέχει τα στοιχεία σύμφωνα με την ονομασία τους και όλα μαζί θα συνδέονται βάση των κλειδιών τους όπως φαίνεται και στο σχήμα της βάσης στην αρχή του report. Όταν αποφασίσαμε να ορίσουμε ως κλειδί των χωρών το **ISO\_Code** συνειδητοποιήσαμε πως κάποιες χώρες μέσα στα αρχεία δεν υπάρχουν στο αρχείο **countries** άρα δεν τους έχει ανατεθεί τιμή κλειδιού. Εφόσον δεν θέλαμε να χαθούν αυτά τα δεδομένα αποφασίσαμε να τους δώσουμε ένα δικό μας **ISO\_Code** που θα ανέρχεται από την μεγαλύτερη τιμή που υπάρχει ήδη και θα αυξάνεται όσο βρίσκουμε χώρες χωρίς **ISO\_Code**.

Τι περιέχουν τα **τροποποιημένα αρχεία**:

**out\_countries.csv**: Περιέχουν μόνο τα columns που μας αφορούν πιο πολύ τα οποία είναι: **ISO\_Code , Display\_Name, ISO, ISO3, FIPS, Region Name, Country\_Status, Developed\_or\_Developing, Population.** Πρακτικά το **countries** αλλά μόνο με τα columns που χρειαζόμαστε.

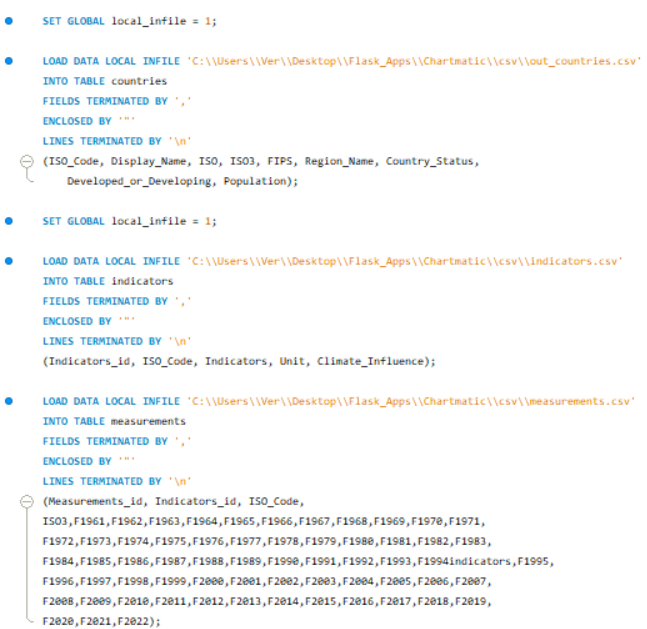
**Indicators.csv** : **Indiactors\_id, ISO\_Code, Indicators, Unit, Climate\_Influence**. Έχουμε βρει για κάθε **ISO\_Code** τα **indicators** που υπάρχουν στα υπόλοιπα αρχεία και τα έχουμε ενσωματώσει στο συγκεκριμένο αρχείο. Η επιλογή και η αντιστοίχιση των δεδομένων σε πρώτο χρόνο έγινε βάση του **ISO3**.

**Measurements.csv :** **Measurements\_id, Indiacotrs\_id, ISO\_Code, ISO\_3, F1961, .. F2022.** Με παρόμοιο τρόπο αντιστοιχήσαμε τα δεδομένα και κρατήσαμε εδώ όλες τις μετρήσεις όλων των χορών για κάθε χρονιά.

ΜΑΪΟΣ 2024

Η λογική του **script** είναι πως διαβάζει τα στοιχεία των χωρών και φτιάχνει το πρώτα το αρχείο countries και το φορτώνει σε μια λίστα. Για κάθε αρχείο παραλείπει την επικεφαλίδα , αναλύει τα δεδομένα για τους δείκτες και τις μετρήσεις και αν δεν βρει **ISO\_Code** για την χώρα, δημιουργεί καινούργιο. Φτιάχνει τις λίστες για **indicators** και για κάθε χρονιά φτιάχνει την λίστα **measurements**. Έπειτα, δημιουργεί ένα **χάρτη ISO3 σε ISO** και με κλιμακούμενο διάβασμα/ επεξεργασία δημιουργεί τις εγγραφές.

Τελευταίο βήμα του ELT είναι να φορτώσουμε τα δεδομένα μας στην βάση. Χρησιμοποιήσαμε την εντολή **LOAD DATA LOCAL INFILE (‘PATH’)**



ΜΑΪΟΣ 2024

**ETL- DIAGRAM**

**Εικόνα που περιέχει κείμενο, στιγμιότυπο οθόνης, διάγραμμα, γραμμή

Περιγραφή που δημιουργήθηκε αυτόματα**

ΜΑΪΟΣ 2024

**USER STORIES**

**CREATE BAR CHARTS**

* AS A: analyst
* -I WANT: to create bar charts for multiple indicators of multiple countries
* SO THAT: to visually draw the data and extract conclusions

**CREATE LINE CHARTS**

* AS A: analyst
* I WANT: to create line charts for multiple indicators of multiple countries
* SO THAT: to visually draw the data and extract conclusions

**CREATE SCATTER PLOTS**

* AS A: analyst
* I WANT: to create scatter plots for two indicators of 1 country
* SO THAT: to visually draw the data and extract conclusions

**FILTER YEARS**

* AS A: analyst
* I WANT: to filter the years of the selected indicators for visualization
* SO THAT: to visualize only the selected year periods

**AGGREGATE YEARS**

* AS A: analyst
* I WANT: to aggregate years into 5-year or 10-year periods
* SO THAT: to calculate the mean score of the visualized data

ΜΑΪΟΣ 2024

**UML USE CASE DIAGRAM**

**Εικόνα που περιέχει σκίτσο/σχέδιο, ζωγραφιά, διάγραμμα, κύκλος

Περιγραφή που δημιουργήθηκε αυτόματα**

ΜΑΪΟΣ 2024

**Για να δημιουργήσουμε Backup στην βάση:**

**Μεταβαίνουμε στο directory : Navigator ->Advanced->Data Export-> Select my\_dbms -> Export Data**

Από το Workbench της MySQL και το section Navigator πηγαίνουμε στο tab Advanced. Από εκεί επιλέγουμε το Data Export και μετά την βάση που θέλουμε να κάνουμε το Backup στην περίπτωση μας την my\_dbms και κάνουμε Export Data. Τα αρχεία και των 3 tables έχουν αποθηκευτεί σε ένα default φάκελο. Τα μεταφέρουμε στον φάκελο του προτζεκτ μας.

ΜΑΪΟΣ 2024

**Architecture Diagram της εφαρμογής μας**

**Εικόνα που περιέχει διάγραμμα, Σχέδιο, τεχνικό σχέδιο, γραμμή

Περιγραφή που δημιουργήθηκε αυτόματα**

* **Csv** : Περιέχει τα αρχικά csv , το script για την διαμόρφωση καινούργιων csv καθώς και την τα διαμορφωμένα csv που θα γίνουν load στην βάση.
* **Charts:** Εκεί υπάρχουν όλα τα scripts για τα charts και ένας φάκελος Images που αποθηκεύονται τοπικά οι εικόνες των διαγραμμάτων.
* **Images:** Εκεί αποθηκεύονται όλες οι εικόνες που γυρνάνε τα chart script.
* **App.py** : Περιέχει τα routes και έχει βασική λειτουργία να ενώσει την βάση και τα scripts των διαγραμμάτων με την html.
* **Templates:** Περιέχει το αρχείο home.html που δίνει στον server όλο το front-end.

ΜΑΪΟΣ 2024