



Σχολή Θετικών Επιστημών και Τεχνολογίας  
Μεταπτυχιακή Εξειδίκευση στα Πληροφοριακά Συστήματα

Διπλωματική Εργασία

Ανάπτυξη αποθήκης δεδομένων για την ανάλυση και διαχείριση  
σεισμολογικών δεδομένων

Γεώργιος Πραμαντιώτης

Επιβλέπουσα καθηγήτρια: Γεωργία Γκαράνη

Πάτρα, Ιούνιος 2023

Η παρούσα εργασία αποτελεί πνευματική ιδιοκτησία του φοιτητή (Γεώργιος Πραμαντιώτης) που την εκπόνησε. Στο πλαίσιο της πολιτικής ανοικτής πρόσβασης ο συγγραφέας/δημιουργός εκχωρεί στο ΕΑΠ, μη αποκλειστική άδεια χρήσης του δικαιώματος αναπαραγωγής, προσαρμογής, δημόσιου δανεισμού, παρουσίασης στο κοινό και ψηφιακής διάχυσής τους διεθνώς, σε ηλεκτρονική μορφή και σε οποιοδήποτε μέσο, για διδακτικούς και ερευνητικούς σκοπούς, άνευ ανταλλάγματος και για όλο το χρόνο διάρκειας των δικαιωμάτων πνευματικής ιδιοκτησίας. Η ανοικτή πρόσβαση στο πλήρες κείμενο για μελέτη και ανάγνωση δεν σημαίνει καθ' οιονδήποτε τρόπο παραχώρηση δικαιωμάτων διανοητικής ιδιοκτησίας του συγγραφέα/δημιουργού ούτε επιτρέπει την αναπαραγωγή, αναδημοσίευση, αντιγραφή, αποθήκευση, πώληση, εμπορική χρήση, μετάδοση, διανομή, έκδοση, εκτέλεση, «μεταφόρτωση» (downloading), «ανάρτηση» (uploading), μετάφραση, τροποποίηση με οποιονδήποτε τρόπο, τμηματικά ή περιληπτικά της εργασίας, χωρίς τη ρητή προηγούμενη έγγραφη συναίνεση του συγγραφέα/δημιουργού. Ο συγγραφέας/δημιουργός διατηρεί το σύνολο των ηθικών και περιουσιακών του δικαιωμάτων.



Ανάπτυξη αποθήκης δεδομένων για την ανάλυση και διαχείριση  
σεισμολογικών δεδομένων

Γεώργιος Πραμαντιώτης

Επιτροπή Επίβλεψης Διπλωματικής Εργασίας

Επιβλέπουσα Καθηγήτρια:

Γεωργία Γκαράνη

«Αναπληρώτρια Καθηγήτρια  
Πανεπιστημίου Θεσσαλίας»

Συν-Επιβλέπων Καθηγητής:

Μιχαήλ Βασιλακόπουλος

«Καθηγητής Πανεπιστημίου Θεσσαλίας»

Πάτρα, Ιούνιος 2023

*«Στην οικογένεια, τους φίλους μου και σε όσους δεν βρίσκονται κοντά μας»*

## Περίληψη

Η επιστήμη της Σεισμολογίας ως κλάδος της Γεωφυσικής έχει αντικείμενο τη μελέτη των σεισμικών φαινομένων χρησιμοποιώντας τα τελευταία χρόνια μεγάλο όγκο δεδομένα, λόγω της αυξανόμενης ικανότητας δημιουργίας και συλλογής πληροφοριών ως αποτέλεσμα μαζικών παρατηρήσεων που έχουν παραδώσει τεράστιες ποσότητες σεισμολογικών δεδομένων. Η συγκέντρωση και ενσωμάτωση πολυάριθμων, ετερογενών πηγών δεδομένων με σκοπό τη δημιουργία μιας κεντρικής, οργανωμένης και ενοποιημένης αποθήκης δεδομένων είναι ζωτικής σημασίας στην αξιοποίηση των δεδομένων αυτών.

Η εργασία αυτή αφορά στο σχεδιασμό και την υλοποίηση μιας αποθήκης δεδομένων με σεισμολογικά δεδομένα με σκοπό την ενσωμάτωση ετερογενών σεισμολογικών δεδομένων και την αποθήκευσή τους, καθώς και την ανάλυση τύπων ερωτημάτων για την εξαγωγή υπολογιστικών και συγκριτικών συμπερασμάτων που θα επιτρέπει στους χρήστες της, ερευνητές και μη, να επεξεργάζονται και να αναλύουν τέτοιου είδους δεδομένα παρέχοντας μια πιο αποτελεσματική αλληλεπίδραση μεταξύ των δεδομένων και των χρηστών. Επιπλέον, εργαλεία όπως μηχανές οπτικής ανάλυσης μπορούν να διευκολύνουν τη δημιουργία διαδραστικών οπτικών αναλυτικών στοιχείων για τη μετατροπή των δεδομένων σε κατανοητά, διαδραστικά γραφικά.

Η συλλογή των δεδομένων θα γίνει με χρήση δυναμικών χαρτών και άλλων πηγών και θα αφορούν διάφορα κριτήρια όπως μέγεθος σεισμού, χρονική περίοδος, εστιακό βάθος, γεωγραφική θέση epicέντρου, κ.ά. αλλά και διάφορες σεισμολογικές παραμέτρους από τους σεισμολογικούς σταθμούς.

### Λέξεις – Κλειδιά

σεισμολογία, σεισμολογικά δεδομένα, αποθήκες δεδομένων, βάσεις δεδομένων, χωροχρονικές βάσεις δεδομένων, οπτικοποίηση δεδομένων

# Data Warehouse Development for Analysis and Management of Seismological Data

Georgios Pramantiotis

## **Abstract**

The science of Seismology as a branch of geophysics has aimed to study seismic phenomena using large amounts of data in recent years due to the increasing ability to generate and collect information as a result of mass observations that have provided huge amounts of seismic data. The aggregation and integration of numerous, heterogeneous data sources into a central, organized and unified data warehouse is crucial for the utilization of this data.

This work concerns the design and implementation of a seismological data warehouse to integrate and store heterogeneous seismological data, as well as the analysis of query types to extract computational and comparative conclusions that will enable its users, researchers and non-researchers, process and analyze such data to enable more efficient interaction between data and users. Additionally, tools such as visual analysis engines can facilitate the creation of interactive visual analysis to transform data into understandable, interactive graphs.

Data collection is carried out using dynamic maps and other sources and concern various criteria such as earthquake size, time period, focus depth, geographic location of the epicenter etc. but also various seismological parameters from the seismological stations.

## **Keywords**

seismology, seismological data, data warehouses, spatiotemporal databases, data visualization

## Περιεχόμενα

Περίληψη .....	v
Abstract .....	vi
Περιεχόμενα .....	vii
Κατάλογος Εικόνων / Σχημάτων.....	ix
Κατάλογος Πινάκων .....	x
Συντομογραφίες & Ακρωνύμια.....	xi
1. Εισαγωγή.....	1
1.1 Στόχοι Διπλωματικής Εργασίας.....	2
1.2 Δομή Διπλωματικής Εργασίας.....	2
2. Σεισμοί.....	3
2.1 Γενικά χαρακτηριστικά σεισμών .....	3
3. Αποθήκες Δεδομένων .....	5
3.1 Σχήματα αποθηκών δεδομένων .....	5
3.1.1 Σχήμα Αστέρα.....	5
3.1.2 Σχήμα Χιονονιφάδας.....	7
3.1.3 Σχήμα Αστερισμού.....	8
3.2 Online Analytical Processing (OLAP).....	9
3.3 Online Transaction Processing (OLTP).....	10
3.4 Εξαγωγή, Μορφοποίηση, Φόρτωση (ETL).....	10
4. Βιβλιογραφική Ανασκόπηση .....	12
5. Ανάπτυξη αποθήκης δεδομένων σεισμών .....	14
5.1 Εργαλεία που χρησιμοποιήθηκαν .....	14
5.2 Διαδικασία ETL.....	15
5.2.1 Τύποι δεδομένων.....	15
5.2.2 Σεισμικά Ρήγματα .....	16
5.2.3 Ζώνες Καταβυθίσεων .....	17
5.2.4 Χώρες .....	18
5.2.5 Περιοχές .....	18
5.2.6 Πόλεις.....	19
5.2.7 Τύποι μεγέθους σεισμών .....	20
5.2.8 Σεισμοί.....	20
5.3 Αρχιτεκτονική αποθήκης δεδομένων.....	23
5.3.1 Date Dimension (dim_dates) .....	24
5.3.2 Time Dimension (dim_times).....	24
5.3.3 Crustal Fault Dimension (dim_faults).....	25
5.3.4 Subduction Discretization Dimension (dim_subductions).....	26
5.3.5 Magnitude Type Dimension (dim_magnitude_types).....	27
5.3.6 City Dimension (dim_cities).....	28
5.3.7 Region Dimension (dim_regions).....	28
5.3.8 Country Dimension (dim_countries).....	29
5.3.9 Fact Earthquake (fact_earthquakes).....	29
5.4 Συναρτήσεις και Εναύσματα (functions & triggers).....	31
5.4.1 Λειτουργία σύνδεσης πόλεων με περιοχές.....	31

5.4.2 Λειτουργία σύνδεσης σεισμών με ζώνες καταβυθίσεων.....	32
5.4.3 Λειτουργία εύρεσης μικρότερης απόστασης μεταξύ επίκεντρου σεισμού και πόλης.....	32
5.4.4 Λειτουργία σύνδεσης σεισμών με σεισμικά ρήγματα.....	33
5.5 Ερωτήματα SQL.....	34
6. Αποτελέσματα Ερωτημάτων SQL.....	35
7. Συμπεράσματα.....	42
Βιβλιογραφία.....	44
Παράρτημα Α: «DDL Αποθήκης δεδομένων» .....	47
Παράρτημα Β: «Ερωτήματα SQL» .....	52



## Κατάλογος Εικόνων / Σχημάτων

Εικόνα 5.1 Πολύγωνο επιλογής δεδομένων ( <a href="https://www.emsc-csem.org/Earthquake/?filter=yes">https://www.emsc-csem.org/Earthquake/?filter=yes</a> ) .....	15
Εικόνα 5.2 Σεισμικά ρήγματα των δεδομένων EFSM20 ( <a href="https://seismofaults.eu/efsm20data">https://seismofaults.eu/efsm20data</a> ) .....	17
Εικόνα 5.3 Ζώνες καταβυθίσεων των δεδομένων EFSM20 ( <a href="https://seismofaults.eu/efsm20data">https://seismofaults.eu/efsm20data</a> ) .....	17
Εικόνα 5.4 Δεδομένα πολυγώνων χωρών countries 2020 ( <a href="https://ec.europa.eu/eurostat/web/gisco/geodata/reference-data/administrative-units-statistical-units/countries#countries20">https://ec.europa.eu/eurostat/web/gisco/geodata/reference-data/administrative-units-statistical-units/countries#countries20</a> ) .....	18
Εικόνα 5.5 Δεδομένα πολυγώνων περιοχών NUTS 2021 level 2 ( <a href="https://ec.europa.eu/eurostat/web/gisco/geodata/reference-data/administrative-units-statistical-units/nuts">https://ec.europa.eu/eurostat/web/gisco/geodata/reference-data/administrative-units-statistical-units/nuts</a> ).....	19
Εικόνα 5.6 Δεδομένα πόλεων ESRI ( <a href="https://hub.arcgis.com/datasets/schools-BE::world-cities/explore">https://hub.arcgis.com/datasets/schools-BE::world-cities/explore</a> ) .....	20
Εικόνα 5.7 1000 από τους 307237 Σεισμούς.....	22
Εικόνα 6.1 Αποτέλεσμα ερωτήματος 1 - σεισμοί εντός ρηγμάτων.....	35
Εικόνα 6.2 Αποτέλεσμα ερωτήματος 2 - σεισμοί εντός ζωνών καταβυθίσεων .....	36
Εικόνα 6.3 Αποτέλεσμα ερωτήματος 3 - 10 μεγαλύτεροι σεισμοί της τελευταίας δεκαετίας κοντά στην Πάτρα .....	36
Εικόνα 6.4 Αποτέλεσμα ερωτήματος 4 - ποσοστό σεισμών ανά χώρα .....	37
Εικόνα 6.5 Αποτέλεσμα ερωτήματος 5 - ποσοστό ρηγμάτων ανά χώρα .....	38
Εικόνα 6.6 Αποτέλεσμα ερωτήματος 6 - 20 πιο σεισμογενείς πόλεις της Ευρώπης ..	38
Εικόνα 6.7 Αποτέλεσμα ερωτήματος 7 - 20 μεγαλύτεροι σεισμοί.....	39
Εικόνα 6.8 Αποτέλεσμα ερωτήματος 9 - σεισμοί ανά έτος.....	40
Εικόνα 6.9 Αποτέλεσμα ερωτήματος 10 - σεισμοί ανά μήνα.....	40
Εικόνα 6.10 Αποτέλεσμα ερωτήματος 11 - σεισμοί ανά μήνα.....	41
Σχήμα 3.1 Σχήμα αστεριού (Kimball & Ross, 2013) .....	6
Σχήμα 3.2 Παιχνίδι ρόλου (Kimball & Ross, 2013) .....	7
Σχήμα 3.3 Σχήμα χιονονιφάδας (Ponniah, 2010) .....	8
Σχήμα 3.4 Σχήμα αστερισμού (Ponniah, 2010).....	9
Σχήμα 5.1 Απεικόνιση σχήματος αποθήκης δεδομένων σεισμών .....	23

## Κατάλογος Πινάκων

Πίνακας 5.1 Περιγραφή τύπων δεδομένων πινάκων .....	16
Πίνακας 5.2 Περιγραφή δεδομένων σεισμών .....	21
Πίνακας 5.3 Περιγραφή πίνακα <i>dim_dates</i> .....	24
Πίνακας 5.4 Περιγραφή πίνακα <i>dim_times</i> .....	25
Πίνακας 5.5 Περιγραφή πίνακα <i>dim_faults</i> .....	26
Πίνακας 5.6 Περιγραφή πίνακα <i>dim_subductions</i> .....	27
Πίνακας 5.7 Περιγραφή πίνακα <i>dim_magnitude_types</i> .....	27
Πίνακας 5.8 Περιγραφή πίνακα <i>dim_cities</i> .....	28
Πίνακας 5.9 Περιγραφή πίνακα <i>dim_regions</i> .....	28
Πίνακας 5.10 Περιγραφή πίνακα <i>dim_countries</i> .....	29
Πίνακας 5.11 Περιγραφή πίνακα <i>fact_earthquakes</i> .....	30
Πίνακας 6.1 Αποτέλεσμα ερωτήματος 8 – ποσοστό σεισμών σε ρήγμα & ποσοστό σεισμών σε ζώνη καταβύθισης .....	39

## Συντομογραφίες & Ακρωνύμια

Ακολουθούν κάποια παραδείγματα:

ΔΕ	Διπλωματική Εργασία
ΕΑΠ	Ελληνικό Ανοικτό Πανεπιστήμιο
ΦΕΚ	Φύλλο Εφημερίδας Κυβέρνησης
OLAP	Online Analytical Processing
OLTP	Online Transaction Processing
ETL	Extract, Transform, Load
DW	Data Warehouse
DB	Database
WHO	World Health Organization
SQL	Structure Query Language
DDL	Data Definition Language
PK	Primary Key
FK	Foreign Key
N	Normal
R	Reverse
LL	Left Lateral
RL	Right Lateral
SS	Strike-slip
DS	Dip-slip

## 1. Εισαγωγή

Οι σεισμοί είναι ένα φυσικό φαινόμενο, το οποίο μπορεί να οδηγήσει από υλικές και οικονομικές ζημιές μέχρι και απώλειες ανθρωπίνων ζωών. Σύμφωνα με την σελίδα του Παγκόσμιου Οργανισμού Υγείας (WHO) όπου αναφέρεται στους σεισμούς, σχεδόν 750.000 θάνατοι προκλήθηκαν παγκοσμίως από σεισμούς την περίοδο 1998-2017 και επηρεάστηκαν περισσότεροι από 125 εκατομμύρια άνθρωποι την περίοδο αυτή (World Health Organization Earthquakes).

Οι σεισμοί είναι μέρος της ζωής εκατομμυρίων ανθρώπων ανά τον κόσμο, επηρεάζοντας την καθημερινότητα και την κοινωνία αυτών. Πολλές χώρες έχουν στραφεί στην ρύθμιση και τη θεσμοθέτηση αντισεισμικών κανονισμών για την ανέγερση κτηρίων (Αμερική, Ευρωπαϊκή Ένωση, Ιαπωνία κτλ.). Ο πρώτος αντισεισμικός κανονισμός στην Ελλάδα ψηφίστηκε με το ΒΔ της 19/26 Φεβρουαρίου 1959 «περί αντισεισμικού κανονισμού οικοδομικών έργων» (ΦΕΚ 36Α/26-2-1959) και βοήθησε στην αντισεισμική προστασία της χώρας την εποχή εκείνη. Έπειτα αναθεωρήθηκε με το ΦΕΚ 239Β/16-4-1984 όπου συμπληρώθηκαν πρόσθετα άρθρα και το 1995 τέθηκε σε εφαρμογή ο Νέος Ελληνικός Αντισεισμικός Κανονισμός (ΝΕΑΚ) με το ΦΕΚ 613Β/12-10-1992. Το 2001 θεσπίστηκε το Ελληνικός Αντισεισμικός Κανονισμός 2000 (ΕΑΚ 2000) με το ΦΕΚ 2184 Β'/20-12-1999. Από το 2014 και έπειτα με την «Έγκριση εφαρμογής και χρήσης Ευρωκωδίκων σε συνδυασμό με τα αντίστοιχα Εθνικά Προσαρτήματα» (ΦΕΚ 1457/2014) μπορεί να χρησιμοποιείται είτε ο Ευρωκώδικας 8 – Αντισεισμικός σχεδιασμός κατασκευών είτε ο ΕΑΚ 2000.

Η καταγραφή και η συστηματική παρακολούθηση των σεισμών μπορεί να μας βοηθήσει στην πιο εύστοχη πρόβλεψη και λήψη αποφάσεων σχετικά με αυτούς. Οι αποθήκες δεδομένων είναι ένα αρκετά εξελιγμένο σύστημα αποθήκευσης και άντλησης δεδομένων, οι οποίες μπορούν να βοηθήσουν στην καταγραφή και την ανάλυση των σεισμών.

Στην διπλωματική αυτή θα αναπτύξουμε μία αποθήκη δεδομένων για σεισμούς, και στην συνέχεια θα αντλήσουμε δεδομένα και στατιστικά με σκοπό την ανάλυση δεδομένων και την εξαγωγή συμπερασμάτων.

## **1.1 Στόχοι Διπλωματικής Εργασίας**

Στόχος της ΔΕ είναι η παρουσίαση ενός μοντέλου αποθήκης δεδομένων η οποία επικεντρώνεται στην διαχείριση και την εξαγωγή σεισμολογικών δεδομένων, μέσω χωρικών, χρονικών και σύνθετων ερωτημάτων, καθώς και η παρουσίαση των αποτελεσμάτων αυτών μέσα από θεματικούς χάρτες και εργαλεία οπτικοποίησης δεδομένων.

## **1.2 Δομή Διπλωματικής Εργασίας**

Η δομή της ΔΕ μετά το κεφάλαιο της Εισαγωγής περιγράφεται παρακάτω. Αρχικά θα ερευνήσουμε τους σεισμούς, τα γενικά χαρακτηριστικά που τους διέπουν, τους τρόπους μέτρησης και τα δεδομένα τους, με σκοπό να καταλάβουμε τις ανάγκες που καλείται να καλύψει η αποθήκη δεδομένων που στοχεύουμε να αναπτύξουμε. Έπειτα θα μιλήσουμε για τις αποθήκες δεδομένων, για τις αρχιτεκτονικές που ακολουθούν και τις διαδικασίες ανάλυσης, επεξεργασίας, εξαγωγής, μετασχηματισμού και φόρτωσης δεδομένων. Στην συνέχεια έχει η βιβλιογραφική ανασκόπηση πάνω στις αποθήκες δεδομένων που έχουν αναπτυχθεί για σεισμολογικά δεδομένα. Μετά την βιβλιογραφική ανασκόπηση σειρά έχει το κεφάλαιο της Εφαρμογής, όπου περιγράφονται όλα τα βήματα που έχουν ακολουθηθεί από την εύρεση των δεδομένων μέχρι και την παρουσίαση αυτών, με αναλυτική περιγραφή όλων των αντικειμένων που έχουν δημιουργηθεί. Τέλος θα παρουσιάσουμε τα αποτελέσματα και θα μιλήσουμε για τα συμπεράσματα τα οποία έχουν προκύψει.

## 2. Σεισμοί

Σύμφωνα με τον Δρούτσο (2000) οι σεισμοί χαρακτηρίζονται από τις δονήσεις του υπεδάφους, που προκαλούνται από τις διαταραχές στη μηχανική ισορροπία του βραχώδους εσωτερικού της γης. Επίσης, μία ακολουθία δονήσεων, που περιλαμβάνει μετασεισμούς και προσεισμούς, σε συνδυασμό με τον κύριο σεισμό, σχηματίζουν μία σεισμική ακολουθία.

Ένα βραχώδες σώμα το οποίο βρίσκεται υπό πίεση, παραμορφώνεται. Όταν το σώμα δεν αντέχει άλλο την παραμόρφωση, σπάει και οι δύο πλευρές γλιστρούν η μία δίπλα στην άλλη. Οι περισσότεροι σεισμοί συμβαίνουν κοντά σε όρια τεκτονικών πλακών, αλλά όχι απαραίτητα πάνω σε ένα όριο, ούτε καν σε προϋπάρχον ρήγμα (Earle, 2015).

### 2.1 Γενικά χαρακτηριστικά σεισμών

Οι σεισμοί σαν φυσικά φαινόμενα, διέπονται από κάποια γενικά χαρακτηριστικά, τα οποία απευθύνονται και χαρακτηρίζουν όλους τους σεισμούς.

Η βασικότερη μονάδα που χρησιμοποιείται για την μέτρηση ενός σεισμού είναι το μέγεθος του (magnitude), το οποίο αντανακλά την ενέργεια που έχει εκλύσει. Αυτό συνήθως υπολογίζεται χρησιμοποιώντας τη λογαριθμική κλίμακα Ρίχτερ, όπως εξηγήθηκε από τους Παπανικολάου και Σιδέρη (2009). Για να προσδιορίσουν το μέγεθος ενός σεισμού, οι επιστήμονες αναλύουν το πλάτος των διαφορετικών σεισμικών κυμάτων και τη χρονική διαφορά μεταξύ της άφιξης των κυμάτων P και S, όπως καταγράφεται από τους σεισμογράφους.

Το μέτρο έντασης ενός σεισμού εκτιμάται εμπειρικά και καθορίζεται από την επίδραση του σε μία δεδομένη τοποθεσία. Κατά την εκτίμηση της έντασης, λαμβάνονται υπόψη πολλοί παράγοντες, όπως το μέγεθος, η διάρκεια και το εστιακό βάθος ενός σεισμού, η επιτάχυνση του εδάφους, οι τύποι πετρωμάτων και η απόσταση από το επίκεντρο του σεισμού μίας περιοχής, η θέση και οι διαστάσεις του ρήγματος που προκάλεσε τον σεισμό, καθώς και τον τύπο και τη σύνθεση των κατασκευών μίας περιοχής (Earle, 2015), (Παπανικολάου & Σιδέρη, 2009). Υπάρχουν διάφορες μέθοδοι που χρησιμοποιούνται για τον προσδιορισμό της έντασης ενός σεισμού, όπως η κλίμακα Mercalli, η κλίμακα MSK και η EMS (European Macroseismic Scale) (Παπανικολάου & Σιδέρη, 2009).

Κατά την γέννηση ενός σεισμού παράγονται κύματα χώρου τα οποία μπορεί να είναι επιμήκη και εγκάρσια. Τα επιμήκη κύματα διαδίδονται με μεγαλύτερη ταχύτητα και ονομάζονται κύματα p (primary), ενώ τα εγκάρσια ονομάζονται κύματα s (secondary) (Δρούτσος, 2000).

Τα επιφανειακά κύματα σχηματίζονται στην επιφάνεια της γης. Αυτά περιλαμβάνουν τα κύματα Rayleigh, όπου τα σημεία στο έδαφος κινούνται με περιστροφικό τρόπο παράλληλα με την κατεύθυνση του κύματος και τα κύματα Love η κίνηση των σημείων είναι κάθετη προς την κατεύθυνση των κυμάτων (Δρούτσος, 2000).

Δύο εξίσου σημαντικοί όροι περιγραφής σεισμών είναι το υπόκεντρο (ή εστία) και το επίκεντρο ενός σεισμού. Το υπόκεντρο είναι το ακριβές σημείο του σεισμού στον χώρο σε συγκεκριμένο βάθος (εστιακό βάθος), και το επίκεντρό του είναι η προβολή του στην επιφάνεια της γης (Earle, 2015).

### 3. Αποθήκες Δεδομένων

Σύμφωνα με τους Han et al. (2012) ως αποθήκη δεδομένων ορίζεται ένα αποθετήριο δεδομένων το οποίο διατηρείται ξεχωριστά από τις λειτουργικές βάσεις δεδομένων ενός οργανισμού, επιτρέπει την ενσωμάτωση διαφόρων συστημάτων εφαρμογών και υποστηρίζει την επεξεργασία πληροφοριών παρέχοντας ιστορικά δεδομένα προς ανάλυση. Μία άλλη προσέγγιση, όπως προτάθηκε από τον Inmon (1996), προτείνει ότι μία αποθήκη δεδομένων είναι μία συλλογή δεδομένων που είναι υποκειμενοστραφής, μεταβαλλόμενη στο χρόνο, ολοκληρωμένη και μη ασταθής. Σκοπός της είναι να συμβάλει στην λήψη αποφάσεων από τη διοίκηση. Επιπλέον, όπως δήλωσε ο Ponniah (2010), η αποθήκη δεδομένων είναι ένα πολύτιμο περιβάλλον πληροφοριών που προσφέρει μια συνολική και ολοκληρωμένη εικόνα της επιχείρησης. Διευκολύνει την πρόσβαση σε τρέχουσες και ιστορικές πληροφορίες για τη λήψη στρατηγικών αποφάσεων, ενώ επιτρέπει επίσης συναλλαγές υποστήριξης αποφάσεων χωρίς παρεμβολές στα λειτουργικά συστήματα. Επιπλέον, προάγει τη συνέπεια στις οργανωτικές πληροφορίες και χρησιμεύει ως μια ευέλικτη και διαδραστική πηγή στρατηγικών πληροφοριών.

#### 3.1 Σχήματα αποθηκών δεδομένων

Σε αυτό το κεφάλαιο γίνεται μία αναφορά στα δημοφιλέστερα σχήματα αποθηκών δεδομένων που συναντώνται στην βιβλιογραφία, στα χαρακτηριστικά τους καθώς και σε περιπτώσεις που έχουν χρησιμοποιηθεί.

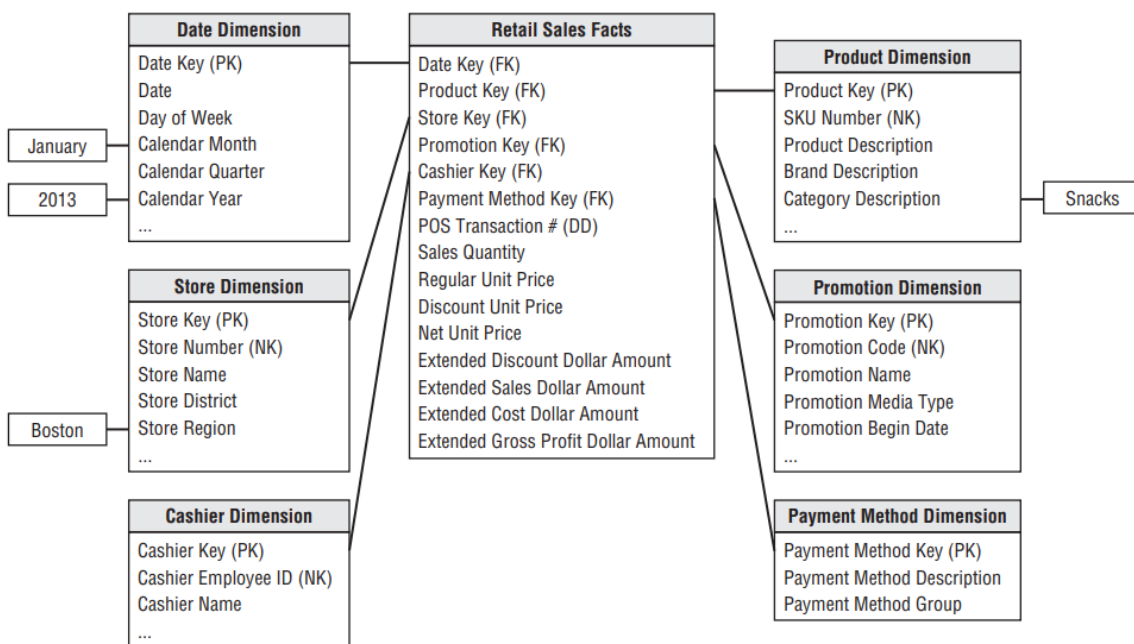
##### 3.1.1 Σχήμα Αστέρα

Τα δεδομένα σε ένα σχήμα αστέρα (βλέπε Σχήμα 1 ) είναι οργανωμένα σε έναν κεντρικό πίνακα του αντικειμένου ενδιαφέροντος (fact table), και σε διάφορους πίνακες διαστάσεις (dimension tables), οι οποίοι δεν περιέχουν υποπίνακες, (Rainardi, 2008). Οι πιο κοινές διαστάσεις είναι η ημερομηνία και η ώρα, αλλά μπορούμε να συναντήσουμε πολλές διαφορετικές ανάλογα με τα δεδομένα που έχουμε αποθηκευμένα μέσα στην αποθήκη δεδομένων.



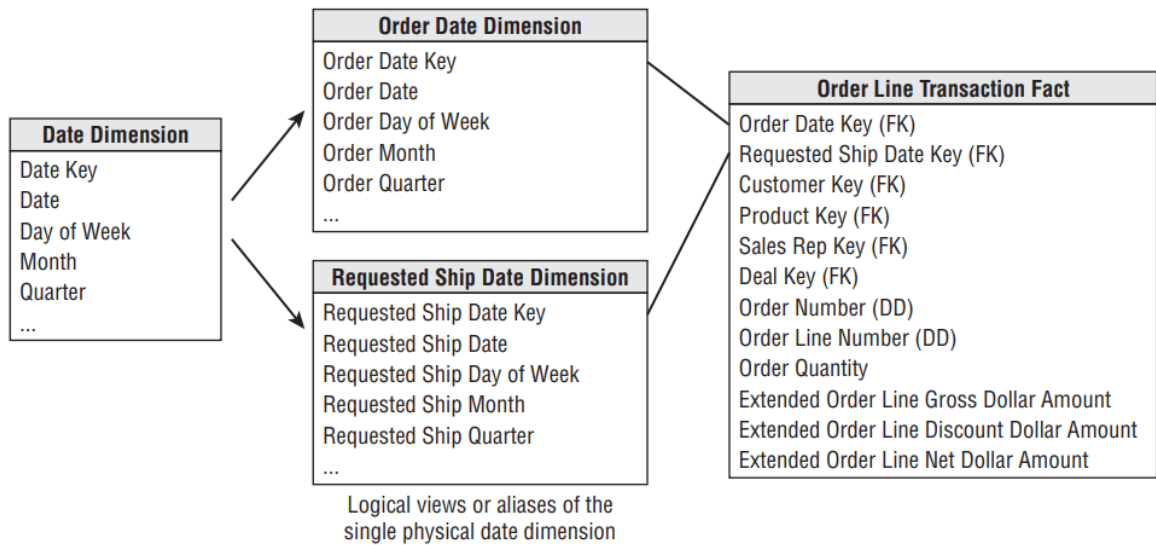
Το πλεονέκτημα του σχήματος αστέρα είναι ότι είναι πιο απλό από το σχήμα χιονονιφάδας διευκολύνοντας έτσι διαδικασίες εξαγωγής, μορφοποίησης και φόρτωσης (ETL) δεδομένων.

Σύμφωνα με τους Kimball & Ross (2013) όπου γίνεται μία πλούσια αναφορά σε περιπτώσεις όπου έχει χρησιμοποιηθεί το σχήμα αστέρα όπως λιανικές πωλήσεις, διαχείριση προμηθειών, διαχείριση παραγγελιών και διαχείριση ανθρωπίνου δυναμικού. Στην περίπτωση του καταστήματος λιανικών πωλήσεων, ο πίνακας γεγονότων αναφέρεται στις πωλήσεις και συνδέεται με τους πίνακες διαστάσεων ημερομηνία, κατάστημα, ταμίας, προϊόν, προώθηση και μέθοδος πληρωμής. Στην συνέχεια παρουσιάζεται η επεκτασιμότητα του παραπάνω σχήματος στην περίπτωση όπου μπορούν να αποθηκευτούν και δεδομένα πελατών με σκοπό την ανάλυση μοτίβων αγορών συναρτήσει γεωγραφικών, δημογραφικών και χαρακτηριστικών συμπεριφοράς πελατών.



Σχήμα 3.1 Σχήμα αστεριού (Kimball & Ross, 2013)

Ένα ακόμα ενδιαφέρον χαρακτηριστικό που παρουσιάζεται είναι το παιχνίδι ρόλων διαστάσεων (βλέπε Σχήμα 2), όπου στο παράδειγμα που παρουσιάζεται από τους Kimball & Ross (2013) μία διάσταση ημερομηνιών δεν συνδέεται απευθείας με τον πίνακα γεγονότων (πίνακας παραγγελιών), αλλά συνδέεται με δύο πίνακες διαστάσεων ημερομηνία παραγγελίας και ημερομηνία αποστολής οι οποίοι με την σειρά τους συνδέονται με τον πίνακα γεγονότων.

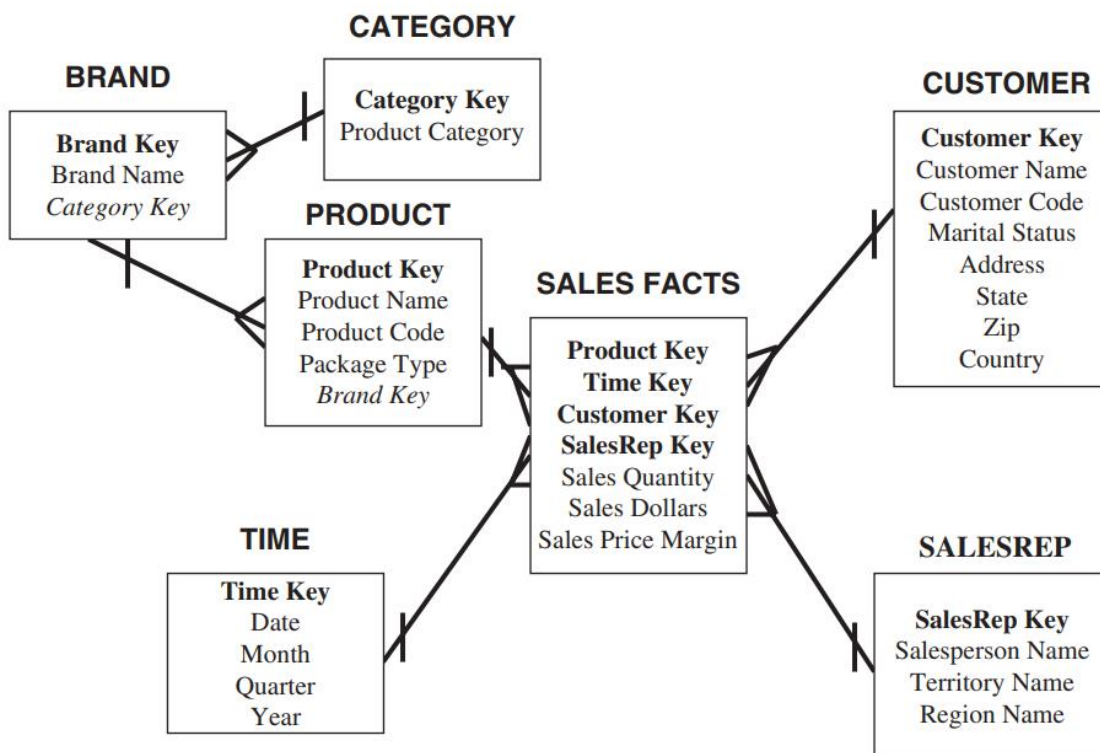


Σχήμα 3.2 Παιχνίδι ρόλου (Kimball & Ross, 2013)

### 3.1.2 Σχήμα Χιονοφιάδας

Το σχήμα χιονοφιάδας (βλέπε Σχήμα 3) είναι παρόμοιο με το star schema με την διαφορά ιεραρχικές συσχετίσεις πινάκων διαστάσεων κανονικοποιούνται και δημιουργούνται υποπίνακες (Kimball & Ross 2013). Ο σκοπός των υποπινάκων είναι η ελαχιστοποίηση περιττών δεδομένων (Rainardi, 2008). Ένα παράδειγμα κανονικοποίησης ενός πίνακα διάστασης δίνεται από τον Ronniah (2010), όπου σε ένα σχήμα αστεριού μία αποθήκης δεδομένων πωλήσεων κανονικοποιεί την διάσταση των προϊόντων και την χωρίζει στις διαστάσεις προϊόν, μάρκα και κατηγορία.

Θετικό στοιχείο του συγκεκριμένου σχήματος είναι ότι έχει λιγότερα περιττά δεδομένα, επομένως καταλαμβάνει λιγότερο χώρο στον δίσκο. Επίσης κανονικοποιημένες δομές είναι ευκολότερο να συντηρηθούν και να ενημερωθούν (Ronniah, 2010).



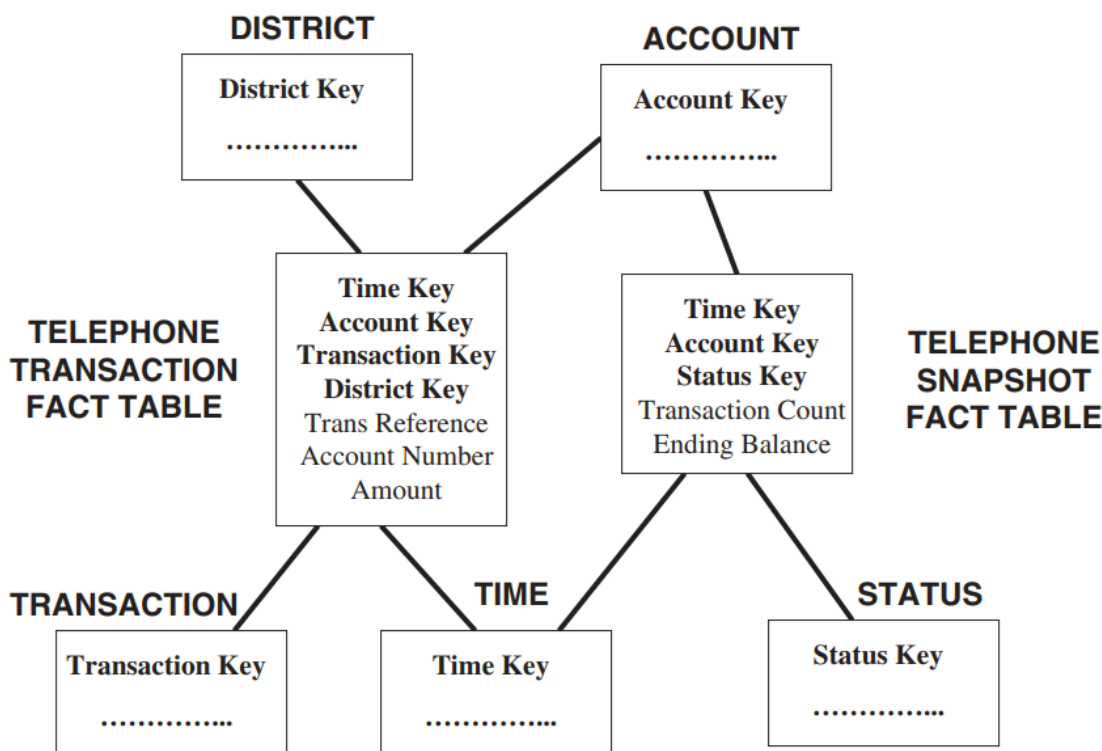
Σχήμα 3.3 Σχήμα χιονονιφάδας (Ponniiah, 2010)

### 3.1.3 Σχήμα Αστερισμού

Σε ένα σχήμα αστερισμού (βλέπε Σχήμα 4) έχουμε δύο ή περισσότερους πίνακες γεγονότων, οι οποίοι συνδέονται σε κοινές διαστάσεις (Rainardi, 2008). Στην ουσία είναι πολλά σχήματα αστέρα ενοποιημένα εξ ου και η ονομασία αστερισμός.

Το πλεονέκτημα του σχήματος αστερισμού είναι ότι μπορεί να μοντελοποιήσει μία επιχείρηση με μεγαλύτερη ακρίβεια χρησιμοποιώντας περισσότερους πίνακες γεγονότων.

Ο Ponniiah (2010) παρουσιάζει μία περίπτωση σχήματος αστερισμού σε μία αποθήκη δεδομένων τηλεφωνικής εταιρίας, όπου υπάρχουν οι πίνακες γεγονότων τηλεφωνικών συναλλαγών και τηλεφωνικών στιγμιотύπων, οι οποίοι έχουν κοινές τις διαστάσεις λογαριασμός και χρόνος και μη κοινές τις διαστάσεις περιοχή, συναλλαγή και κατάσταση.



Σχήμα 3.4 Σχήμα αστερισμού (Ponniah, 2010)

### 3.2 Online Analytical Processing (OLAP)

Σύμφωνα με τον Rainardi (2008), OLAP είναι η δραστηριότητα της διαδραστικής ανάλυσης δεδομένων και συναλλαγών μίας αποθήκης δεδομένων για τη λήψη τακτικών και στρατηγικών αποφάσεων μίας επιχείρησης. Στο βιβλίο των Han et al. (2012), γίνεται μία εκτενής ανάλυση γύρω από την OLAP και τις κοινές της διαδικασίες. Τα συστήματα OLAP βασίζονται σε πολυδιάστατα μοντέλα δεδομένων, όπως τα σχήματα αστέρα, χιονονιφάδας και αστερισμού. Σκοπός τους είναι κυρίως η άντληση πληροφορίας, επομένως επικεντρώνονται κυρίως στην ανάγνωση και την ανάλυση των δεδομένων, όχι μόνο τωρινών αλλά και ιστορικών. Απευθύνονται εργαζόμενους ενός γνωστικού αντικειμένου και είναι σχεδιασμένα για περίπλοκα ερωτήματα.

Κοινές διαδικασίες OLAP είναι οι roll-up, drill-down, slice and dice, pivot. Η διαδικασία roll-up είναι μία διαδικασία άθροισης σε έναν κύβο δεδομένων ανεβαίνοντας ιεραρχικά και οδηγώντας σε πιο γενικευμένη ανάλυση όπως π.χ. το άθροισμα των σεισμών ανά έτος.

Αντιθέτως η διαδικασία drill-down οδηγεί σε πιο λεπτομερή ανάλυση βάσει των στοιχείων μιας διάστασης, όπως π.χ. το άθροισμα των σεισμών κοντά σε μία πόλη. Η διαδικασία slice and dice εστιάζει σε επιλογή εύρους μιας διάστασης πράγμα που οδηγεί σε έναν μικρότερο κύβο δεδομένων, όπως π.χ. το άθροισμα των σεισμών μίας χώρας από το 2010 μέχρι το 2020. Τέλος η διαδικασία pivot περιστρέφει τα δεδομένα (γραμμές και στήλες) για την εναλλακτική παρουσίαση των δεδομένων.

### 3.3 Online Transaction Processing (OLTP)

Το Online Transaction Processing (OLTP) είναι ένα σύστημα το οποίο έχει σαν σκοπό την σύλληψη και την αποθήκευση των συναλλαγών μίας επιχείρησης (Rainardi, 2008). Τα συστήματα OLTP καλύπτουν το μεγαλύτερο εύρος των καθημερινών διαδικασιών ενός οργανισμού, όπως αγοραπωλησίες, λογιστικά αρχαιοθετήσεις κλπ. (Han et al., 2012). Ένα OLTP σύστημα επικεντρώνεται στις συναλλαγές, για καθημερινή χρήση κυρίως από διαχειριστές και επαγγελματίες βάσεων δεδομένων. Είναι σχεδιασμένο για σύντομες και απλές συναλλαγές με δικαιώματα ανάγνωσης και εγγραφής κυρίως για εισαγωγή δεδομένων και επικεντρώνεται κυρίως σε τωρινά δεδομένα (Han et al., 2012).

### 3.4 Εξαγωγή, Μορφοποίηση, Φόρτωση (ETL)

Τα αρχικά ETL αναφέρονται στους όρους Extract, Transform και Load, και είναι η διαδικασία της ανάκτησης και του μετασχηματισμού των δεδομένων από ένα σύστημα, και η φόρτωσή τους σε μία αποθήκη δεδομένων (Rainardi, 2008). Η διαδικασία ETL και σύμφωνα με τον Han et al. (2010) περιλαμβάνει τις λειτουργίες εξαγωγής δεδομένων, καθαρισμού δεδομένων, μετασχηματισμού δεδομένων, φόρτωσης και ανανέωσης δεδομένων.

Η λειτουργία της εξαγωγής συλλέγει τα δεδομένα από διάφορες πηγές (Han et al., 2010) και σύμφωνα με τους Kimball & Ross (2013), η λειτουργία της εξαγωγής σημαίνει την ανάγνωση και κατανόηση των πηγών δεδομένων και η εισαγωγή τους στο σύστημα ETL για περαιτέρω χειραγώγηση. Η λειτουργία του καθαρισμού έχει σαν σκοπό τον εντοπισμό

λαθών στα δεδομένα και της αναδόμησης τους όπου αυτό είναι εφικτό, μπορεί να περιλαμβάνει διορθώσεις τυπογραφικών λαθών, καθαρισμό διπλότυπων δεδομένων και αντιμετώπιση στοιχείων που λείπουν (Han et al., 2012 Kimball & Ross 2013). Ο μετασχηματισμός των δεδομένων μετατρέπει τα δεδομένα στην μορφή της αποθήκης δεδομένων και μαζί με την λειτουργία καθαρισμού εστιάζουν στην βελτίωση της ποιότητας των δεδομένων. Η λειτουργία της φόρτωσης δεδομένων εστιάζει στον υπολογισμό views μίας βάσης, δημιουργία δεικτών και ελέγχου ακεραιότητας των δεδομένων. Τέλος η λειτουργία της ανανέωσης, μεταδίδει ενημερώσεις δεδομένων από τις πηγές στην αποθήκη δεδομένων (Han et al., 2010).

## 4. Βιβλιογραφική Ανασκόπηση

Εντύπωση προκαλεί το γεγονός ότι, παρ' όλο που και οι σεισμοί και οι αποθήκες δεδομένων είναι δύο ευρέως ερευνημένα αντικείμενα, η ανάπτυξη αποθήκης δεδομένων για σεισμολογικά δεδομένα δεν είναι τόσο δημοφιλής.

Στο άρθρο των Marketos et al. (2008), *Seismological Data Warehousing and Mining: A Survey*, παρουσιάζεται η αρχιτεκτονική μιας χωρικής αποθήκης δεδομένων ως ένα σύστημα διαχείρισης και εξόρυξης σεισμολογικών δεδομένων, εστιάζοντας στην αποτελεσματική αναζήτηση δεδομένων, στην ανάλυση OLAP καθώς και την εξόρυξη δεδομένων.

Το σχήμα που επιλέχθηκε είναι αυτό της χιονονιφάδας, με την διάσταση Time να περιλαμβάνει πεδία όπως δεκαετία, έτος, τετράμηνο, μήνας, μέρα και ώρα. Επίσης, η πληροφορία για το μέγεθος, την ένταση και το βάθος των σεισμών, αποθηκεύεται σε ξεχωριστή διάσταση για κάθε παράμετρο (Magnitude, Intensity και Depth), υπό την μορφή διαστημάτων. Προσδιορίζονται δηλαδή με ένα πεδίο κλειδί, ένα πεδίο 'from' και ένα πεδίο 'to'. Η διάσταση Geography αναφέρεται στο πεδίο και σε μία ακόμη διάσταση «Geology», όπου υπάρχει πληροφορία για την γεωλογία μία περιοχής και είναι το σημείο στο σχήμα της αποθήκης δεδομένων το οποίο το κάνει να είναι χιονονιφάδα. Ο πίνακας των γεγονότων αναφέρεται αριθμό σεισμικών γεγονότων ανά συγκεκριμένο τύπο σεισμών, με πεδία τα κλειδιά των παραπάνω διαστάσεων, τον αριθμό των σεισμών και τιμές για την ελάχιστη, την μέση και την μέγιστη τιμή μεγέθους, έντασης και βάθους σεισμού.

Ένα παράδειγμα των ερωτημάτων για την παραπάνω αποθήκη δεδομένων είναι η εύρεση των δέκα σεισμών, που πραγματοποιήθηκαν τους τέσσερις τελευταίους μήνες και είναι κοντά σε μία δεδομένη τοποθεσία. Ένα ακόμα παράδειγμα ήταν η εύρεση όλων των σεισμών μεγέθους  $M > 5$ , που συνέβησαν σε μία συγκεκριμένη περιοχή, τους τελευταίους τέσσερις μήνες. Ένα τελευταίο παράδειγμα που παρουσιάζεται είναι η εύρεση των πέντε μεγαλύτερων σεισμών που συνέβησαν σε ακτίνα μικρότερη των 100 χιλιομέτρων από πόλεις με πληθυσμό πάνω από ένα εκατομμύριο κατά τον 20<sup>ο</sup> αιώνα.

Μία παραμφερής προσέγγιση παρουσιάζεται από τους Özcan & Peker (2021), όπου αναπτύσσεται μία αποθήκη δεδομένων με σκοπό την εκτίμηση κινδύνου σεισμού στην περίπτωση κτιρίων υγείας. Η αποθήκη δεδομένων έχει το σχήμα αστερισμού με δύο πίνακες

γεγονότων (Fact Risk Assessment και Fact Facility Negative Parameter Evaluation) οι οποίοι μοιράζονται κοινές διαστάσεις.



## 5. Ανάπτυξη αποθήκης δεδομένων σεισμών

Στο κεφάλαιο αυτό μιλάμε για την εφαρμογή και όλα τα βήματα που ακολουθήθηκαν από την συλλογή των δεδομένων την επεξεργασία αυτών, την ανάπτυξη της αποθήκης δεδομένων και την ανάπτυξη των λειτουργιών και των εναυσμάτων που κρίθηκαν αναγκαία για την διασφάλιση της ακεραιότητας των δεδομένων.

### 5.1 Εργαλεία που χρησιμοποιήθηκαν

Για την ανάπτυξη της αποθήκης δεδομένων καθώς και για την επεξεργασία και φόρτωση των δεδομένων χρησιμοποιήθηκαν στο σύνολο τους τα παρακάτω εργαλεία:

- Λειτουργικό Σύστημα Linux Ubuntu 23.04
- PostgreSQL έκδοση 15.3
- pgAdmin4 έκδοση 7.3
- PostGIS έκδοση 3.3
- Python έκδοση 3.11.2
- QGIS έκδοση 3.28.7

Η PostgreSQL και με την επέκταση PostGIS είναι ανοικτού κώδικα και από τις πιο δημοφιλείς βάσεις δεδομένων για την φιλοξενία και ανάλυση γεωχωρικών δεδομένων όπως στην παρούσα περίπτωση. Με πρώτη έκδοση το 2001 έχει 22 έτη συνεχόμενης ενημέρωσης και υποστήριξης γεωχωρικών δεδομένων καθίσταται η πλέον κατάλληλη για την παρούσα αποθήκη δεδομένων. Το pgAdmin είναι το επίσημο πρόγραμμα επεξεργασίας PostgreSQL βάσεων δεδομένων. Η γλώσσα προγραμματισμού python είναι από τις πιο δημοφιλείς γλώσσες υψηλού επιπέδου και θα δώσει πολλές και εύκολες λύσεις στην διαδικασία ETL. Το λογισμικό QGIS είναι το δημοφιλέστερο λογισμικό GIS ανοικτού κώδικα, με άπειρες λειτουργίες και δυνατότητες ανάλυσης και επεξεργασίας γεωχωρικών δεδομένων.

## 5.2 Διαδικασία ETL

Τα δεδομένα που χρησιμοποιήθηκαν στην ανάπτυξη της αποθήκης δεδομένων είναι σεισμοί από το 2004 ως το 2023, σεισμικά ρήγματα, ζώνες καταβυθίσεων, πόλεις, περιοχές και χώρες. Τα δεδομένα της διπλωματικής αφορούν σεισμούς στην Ευρώπη, επομένως δημιουργήθηκε ένα πολύγωνο με συντεταγμένες γεωγραφικό πλάτος από 30 ως 80 μοίρες και γεωγραφικό μήκος από -20 ως 60 μοίρες, το οποίο καλύπτει όλη την Ευρώπη (βλέπε Εικόνα 5). Στα επόμενα κεφάλαια παρουσιάζονται αναλυτικότερα τα δεδομένα και η επεξεργασία τους μέχρι την εισαγωγή τους στην αποθήκη δεδομένων.

Εικόνα 5.1 Πολύγωνο επιλογής δεδομένων (<https://www.emsc-csem.org/Earthquake/?filter=yes>)

### 5.2.1 Τύποι δεδομένων

Η παρούσα αποθήκη δεδομένων είναι μία χωρο-χρονική αποθήκη δεδομένων, δηλαδή φιλοξενεί δεδομένα τα οποία έχουν χρονική και γεωμετρική πληροφορία. Οι τύποι δεδομένων (βλέπε Πίνακας 5.1) γεωμετρικής πληροφορίας που χρησιμοποιήθηκαν στην αποθήκη δεδομένων φαίνονται παρακάτω:

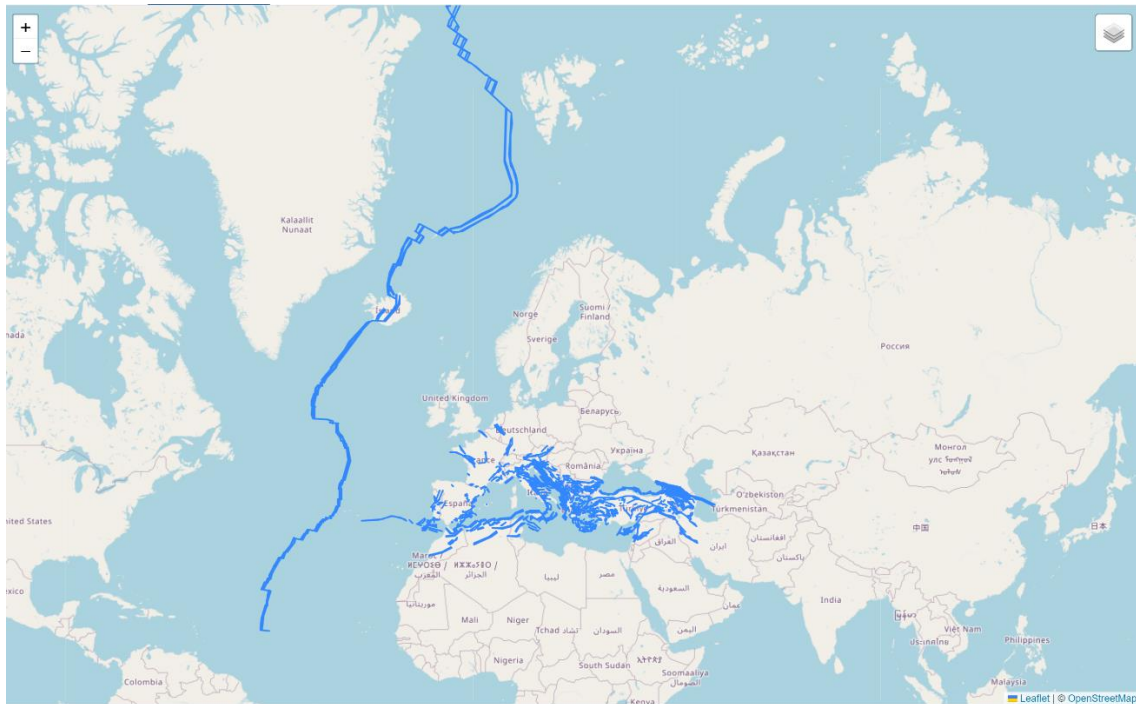
Όνομα πίνακα	Τύπος δεδομένων	Σύστημα Αναφοράς	Περιγραφή
fact_earthquakes	Point	4326	Σημείο επίκεντρου σεισμού
dim_faults	Polygon	4326	Πολύγωνο σεισμικού ρήγματος
dim_subductions	Polygon	4326	Πολύγωνο ζώνης καταβύθισης
dim_countries	MultiPolygon	4326	Πολύγωνο εδάφους χώρας
dim_regions	MultiPolygon	4326	Πολύγωνο εδάφους περιοχής
dim_cities	Point	4326	Σημείο Πόλης

**Πίνακας 5.1 Περιγραφή τύπων δεδομένων πινάκων**

Όλα τα δεδομένα γεωμετρικής πληροφορίας πρέπει να είναι στο ίδιο σύστημα αναφοράς ούτως ώστε να μπορούν τα ερωτήματα να εκτελεστούν ορθά. Το σύστημα αναφοράς θα πρέπει να καλύπτει όλη την Ευρώπη και έτσι επιλέχθηκε το WGS '84 (EPSG: 4326). Τα σημειακά δεδομένα (point) αποτελούνται μόνο από δύο συντεταγμένες (γεωγραφικό μήκος και γεωγραφικό πλάτος). Τα πολυγωνικά (polygon) δεδομένα περιλαμβάνουν την επιφάνεια ενός κλειστού πολυγώνου. Τα δεδομένα multi-polygon μπορούν να περιέχουν περισσότερα τους ενός πολυγώνου ανά πλειάδα στην αποθήκη δεδομένων και επίσης μπορούν να περιέχουν σπές, όπως αυτά αναφέρονται στο επίσημο εγχειρίδιο της (Postgis Documentation, Geometries). Τα multi-polygon δεδομένα είναι κατάλληλα για δεδομένα χωρών οι οποίες περιέχουν νησιά.

### 5.2.2 Σεισμικά Ρήγματα

Δεδομένα για τα σεισμικά ρήγματα είναι μέρος των EMSF 20 (βλέπε Εικόνα 5.2) και ανακτήθηκαν από την Ευρωπαϊκή Βάση Δεδομένων Σεισμικών Ρηγμάτων (European Databases of Seismic Faults, σε μορφή γεωχωρικών δεδομένων με περιγραφική πληροφορία (Basili R et al. 2022). Αφότου επιλέχθηκαν τα δεδομένα τα οποία είναι εντός του πολυγώνου που τέθηκε νωρίτερα, εισήχθησαν μέσα στην αποθήκη δεδομένων στον πίνακα dim\_faults συνολικά 1248 σεισμικά ρήγματα.



Εικόνα 5.2 Σεισμικά ρήγματα των δεδομένων EFSM20 (<https://seismofaults.eu/efsm20data>)

### 5.2.3 Ζώνες Καταβυθίσεων

Τα δεδομένα των ζωνών καταβυθίσεων (βλέπε Εικόνα 5.3) είναι μέρος των EFSM20 της Ευρωπαϊκής Βάσης Δεδομένων Σεισμικών Ρηγμάτων όπως και τα σεισμικά ρήγματα. Αφορούν τμήματα των τεκτονικών πλακών τα οποία βυθίζονται κάτω από διπλανές τους. Το φαινόμενο αυτό παρατηρείται στην Μεσόγειο και συγκεκριμένα δυτικά του Γιβραλτάρ, ανατολικά της Σικελίας και στη ζώνη που ξεκινάει νότια της Πελοποννήσου, συνεχίζει νότια της Κρήτης και φτάνει μέχρι νότια της Κύπρου, όπως στην παρακάτω εικόνα:



Εικόνα 5.3 Ζώνες καταβυθίσεων των δεδομένων EFSM20 (<https://seismofaults.eu/efsm20data>)

#### 5.2.4 Χώρες

Τα δεδομένα διοικητικών ορίων των χωρών (βλέπε Εικόνα 5.4) ανακτήθηκαν από τον ιστότοπο της ευρωπαϊκής στατιστικής υπηρεσίας (Eurostat) και αφορούν το πακέτο δεδομένων Countries 2020. Με την χρήση του προγράμματος QGIS έκδοση 3.28.7 επιλέχθηκαν μόνο οι χώρες οι οποίες τέμνονται από το πολύγωνο που τέθηκε στην αρχή. Οι χώρες αυτές φαίνονται στην επόμενη εικόνα:



**Εικόνα 5.4 Δεδομένα πολυγώνων χωρών countries 2020**  
(<https://ec.europa.eu/eurostat/web/gisco/geodata/reference-data/administrative-units-statistical-units/countries#countries20>)

Η περιγραφική πληροφορία που χρησιμοποιήθηκε από τα παραπάνω δεδομένα είναι μόνο ο κωδικός χώρας (country\_id) και η ονομασία χώρας (country\_name).

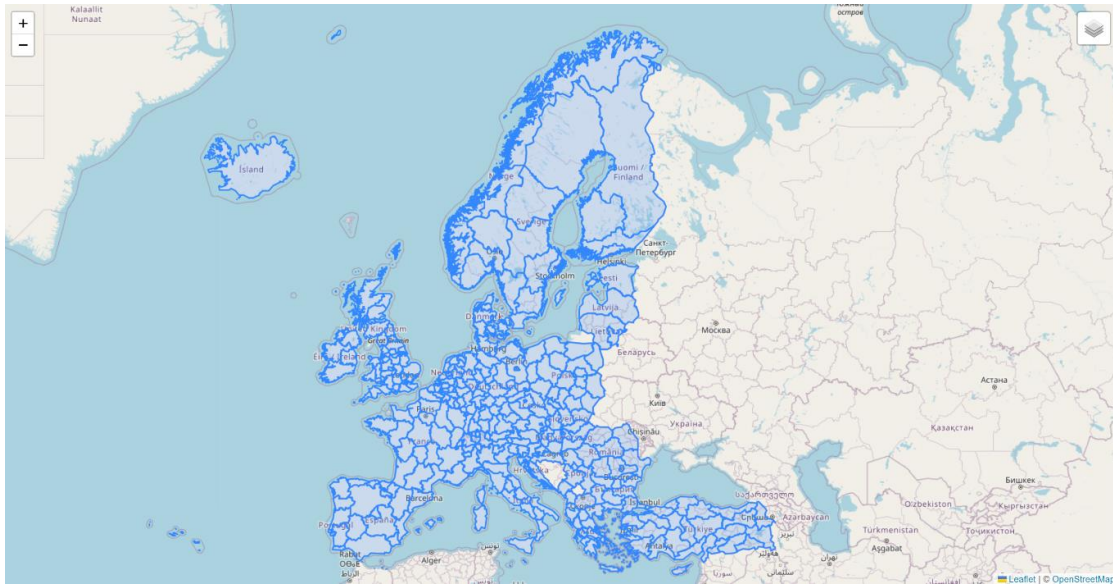
#### 5.2.5 Περιοχές

Τα δεδομένα των περιοχών των ευρωπαϊκών χωρών (βλέπε Εικόνα 5.5) αντλήθηκαν από τον ιστότοπο της ευρωπαϊκής στατιστικής υπηρεσίας (Eurostat) και αφορούν το πακέτο δεδομένων NUTS 2021 level 2. Η ταξινόμηση αυτή (NUTS 2021) έχει ισχύει από την 1<sup>η</sup> Ιανουαρίου 2021 και περιέχει 92 περιοχές στο NUTS 1, 242 περιφέρειες στο NUTS 2 και 1166 περιοχές στο επίπεδο NUTS 3.

- NUTS level 1: κύριες κοινωνικοοικονομικές περιοχές
- NUTS level 2: βασικές περιφέρειες, εφαρμογής περιφερειακών πολιτικών
- NUTS level 3: μικρότερες περιοχές



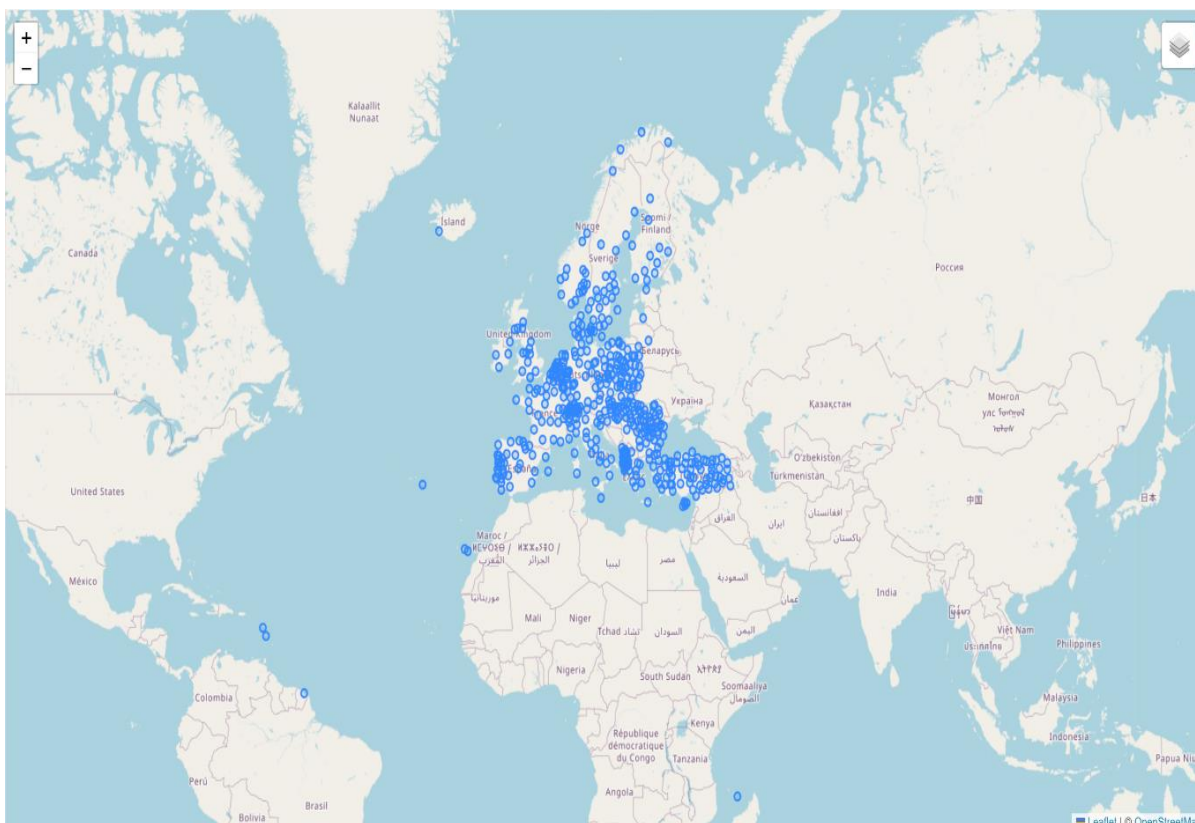
Τα δεδομένα που επιλέχθηκαν είναι οι περιοχές NUTS 2. Η περιγραφική πληροφορία που επιλέχθηκε από το σύνολο των δεδομένων είναι ο κωδικός περιοχής (region\_id), ο κωδικός χώρας (country\_id), ο οποίος έρχεται σε συσχέτιση με τον κωδικό χώρας των δεδομένων countries 2020 και η ονομασία της περιοχής (region\_name).



**Εικόνα 5.5** Δεδομένα πολυγώνων περιοχών NUTS 2021 level 2  
(<https://ec.europa.eu/eurostat/web/gisco/geodata/reference-data/administrative-units-statistical-units/nuts>)

### 5.2.6 Πόλεις

Τα δεδομένα των πόλεων των χωρών (βλέπε Εικόνα 5.6) ανακτήθηκαν από την ESRI (World Cities) και περιλαμβάνουν τις τοποθεσίες των μεγάλων πόλεων του κόσμου. Οι πόλεις περιλαμβάνουν εθνικές πρωτεύουσες, πρωτεύουσες επαρχιών, μεγάλα πληθυσμιακά κέντρα και πόλεις ορόσημα. Το επίπεδο ήταν σε μορφή GEOJSON και για να επιλεγθούν μόνο οι πόλεις της Ευρώπης έγινε χρήση του λογισμικού QGIS έκδοσης 3.28.7, όπου επιλέχθηκαν μόνο οι πόλεις οι οποίες τέμνονται με τα πολύγωνα των περιοχών και φαίνονται στην παρακάτω εικόνα:



Εικόνα 5.6 Δεδομένα πόλεων ESRI (<https://hub.arcgis.com/datasets/schools-BE::world-cities/explore>)

Η περιγραφική πληροφορία που επιλέχθηκε από αυτά τα δεδομένα είναι μόνο η ονομασία της πόλης (city\_name).

### 5.2.7 Τύποι μεγέθους σεισμών

Δεδομένα σχετικά με τους τύπους μεγεθών συμπληρώθηκαν σύμφωνα με τους τύπους που δίνονται από το Γεωλογικό Ινστιτούτο Ηνωμένων Πολιτειών (USGS), τον Bormann, P. (2011) και Bormann et al. (2013).

### 5.2.8 Σεισμοί

Τα δεδομένα των σεισμών (βλέπε Εικόνα 5.7) πάρθηκαν από τον ιστότοπο του Σεισμολογικού Κέντρου Ευρώπης-Μεσογείου (European-Mediterranean Seismological Centre). Επιλέχθηκαν όλοι οι σεισμοί από 2004-10-01 μέχρι 2023-05-31, οι οποίοι βρίσκονται μέσα στο πολύγωνο το οποίο θέσαμε νωρίτερα. Τα δεδομένα κατέβηκαν σε αρχεία csv ανά 5000 εγγραφές, αφορούν 307237 σεισμούς και οι στήλες τους φαίνονται στον Πίνακα 5.2.

Όνομα στήλης	Μονάδα	Περιγραφή
Date	UTC	Ημερομηνία σεισμού
Time	UTC	Ώρα σεισμού
Latitude	degrees	Γεωγραφικό πλάτος επικέντρου σεισμού στο σύστημα αναφοράς WGS '84 (EPSG:4326)
Longitude	degrees	Γεωγραφικό μήκος επικέντρου σεισμού στο σύστημα αναφοράς WGS '84 (EPSG:4326)
Depth	km	Βάθος υπόκεντρου σε χιλιόμετρα
Magnitude Type		Τύπος μεγέθους σεισμού
Magnitude	Richter	Μέγεθος σεισμού
Region Name		Περιοχή και Χώρα σεισμού
Last update	UTC	Ημερομηνία και ώρα τελευταίας ενημέρωσης

**Πίνακας 5.2 Περιγραφή δεδομένων σεισμών**

Από τα παραπάνω δεδομένα αφαιρέθηκαν τα πεδία Region Name και Last Update. Το πεδίο Date μετασχηματίστηκε με την συνάρτηση replace() από τη μορφή YYYY-MM-DD στην μορφή YYYYMMDD για να ταιριάζει με το πρωτεύον κλειδί της διάστασης των ημερομηνιών (dim\_dates).

```
date_id = Date.replace("-", "")
```

#### **Κώδικας 1 Μετασχηματισμός μορφής ημερομηνίας σε python**

Ομοίως το πεδίο Time μετασχηματίστηκε από HH:MM:SS σε HHMMSS για να ταιριάζει με το πρωτεύον κλειδί της διάστασης των ωρών (dim\_times).

```
time_id = Time.replace(":", "")
```

#### **Κώδικας 2 Μετασχηματισμός μορφής ώρας σε python**

Τα πεδία Latitude και Longitude αναφέρονται στο γεωγραφικό πλάτος και στο γεωγραφικό μήκος αντίστοιχα του κάθε σημείου και χρησιμοποιήθηκαν για τον υπολογισμό της γεωμετρίας του επίκεντρου του σεισμού με τις συναρτήσεις ST\_SetSRID η οποία θέτει το σύστημα αναφοράς της γεωμετρίας και την ST\_MakePoint η οποία δημιουργεί την γεωμετρία στον πίνακα fact\_earthquakes.



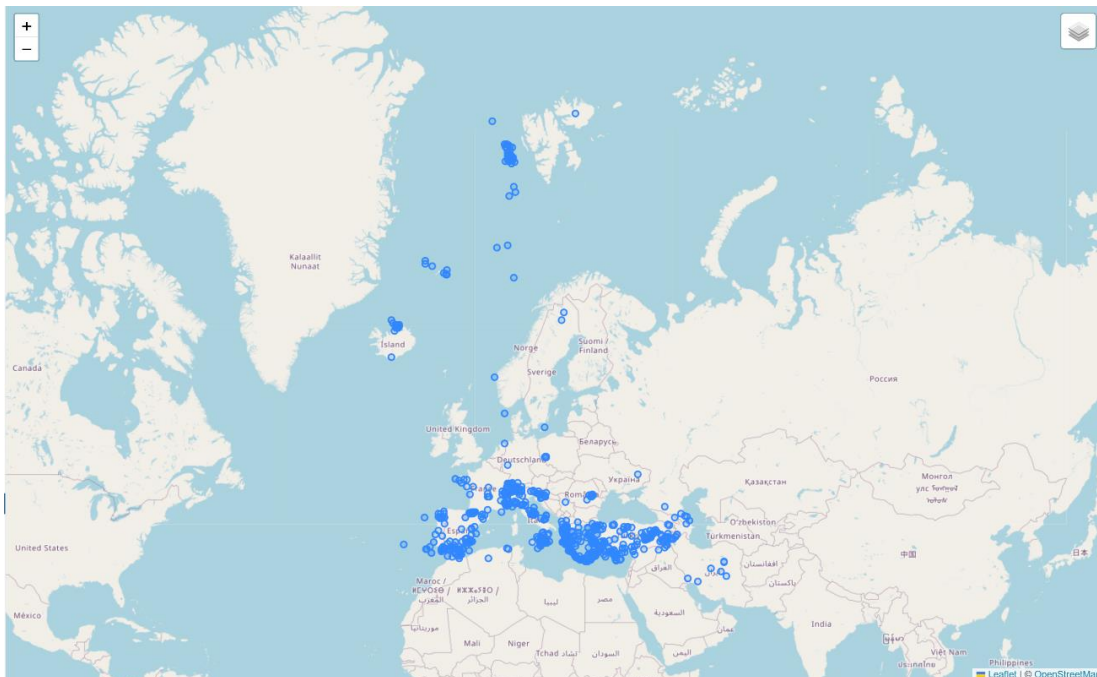
```
ST_SetSRID(ST_MakePoint(Longitude, Latitude), 4326)
```

### Κώδικας 3 Δημιουργία γεωμετρίας από συντεταγμένες σε sql

Επίσης κρίθηκε αναγκαίο να γίνει καθαρισμός δεδομένων στο πεδίο Magnitude Type έτσι ώστε όλες του οι τιμές να ταιριάζουν με τα πρωτεύοντα κλειδιά του πίνακα `dim_magnitude_types`. Πιο συγκεκριμένα καθαρισμός δεδομένων έγινε με το παρακάτω κομμάτι κώδικα, με το οποίο ορίζονται κενές τιμές (NULL) και αντικαθίστανται οι τιμές “MD”, ”M ”, “ M”, “MC”, “ml” με τις τιμές “Md”, “M”, “M”, “Mc” και “ML” αντίστοιχα.

```
if pd.isna(row[6]):
    magnitude_type = None
elif row[6] == "MD":
    magnitude_type = row[6].replace("MD", "Md")
elif row[6] == "M ":
    magnitude_type = row[6].replace("M ", "M")
elif row[6] == " M":
    magnitude_type = row[6].replace(" M", "M")
elif row[6] == "MC":
    magnitude_type = row[6].replace("MC", "Mc")
elif row[6] == "ml":
    magnitude_type = row[6].replace("ml", "ML")
else:
    magnitude_type = row[6]
```

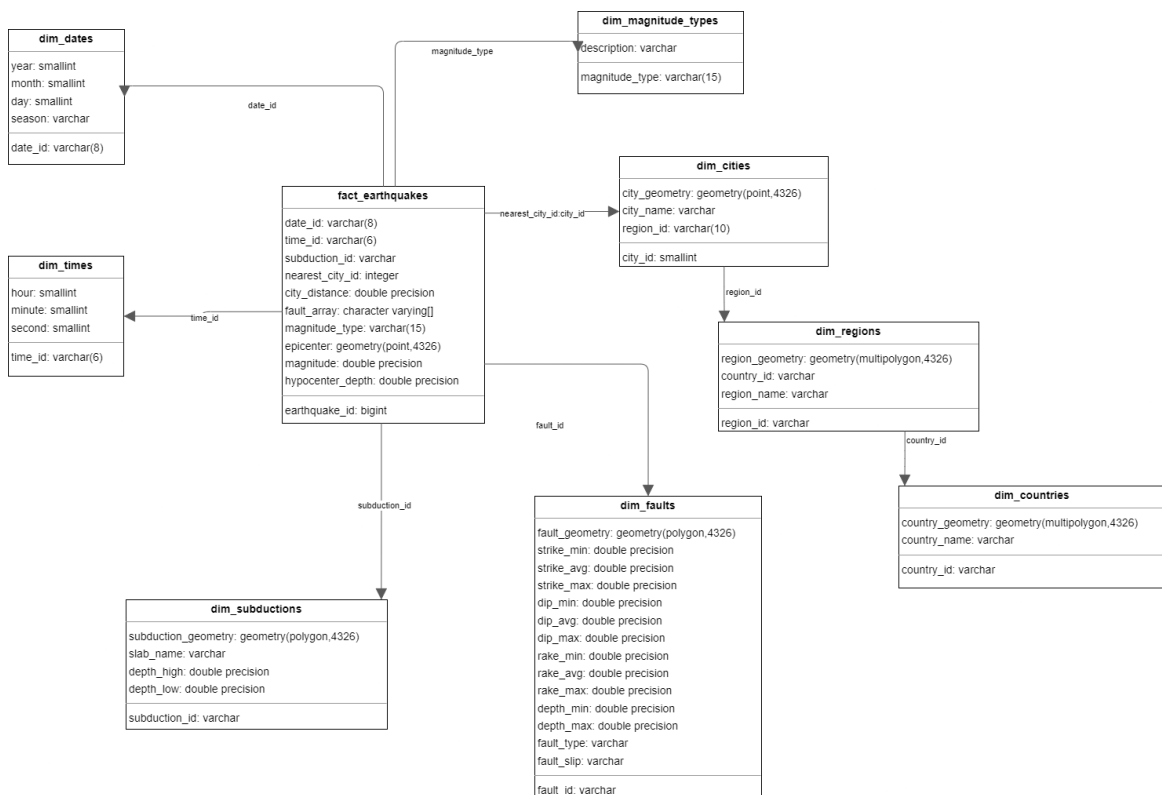
### Κώδικας 4 Καθαρισμός τιμών magnitude\_type σε python



Εικόνα 5.7 1000 από τους 307237 Σεισμούς

### 5.3 Αρχιτεκτονική αποθήκης δεδομένων

Το σχήμα το οποίο επιλέχθηκε για την υλοποίηση της αποθήκης δεδομένων είναι το σχήμα της χιονονιφάδας (βλέπε Σχήμα 5.1). Στην περίπτωση των σεισμών έχουμε μόνο έναν fact πίνακα επομένως θα έπρεπε να επιλέξουμε ανάμεσα σε σχήμα αστεριού και χιονονιφάδας. Στην περίπτωση της διάστασης πόλεων (dim\_city), αποφασίσαμε να εμβαθύνουμε περισσότερο σε πληροφορία, θέτοντας και τους πίνακες dim\_region και dim\_country, που αναφέρονται σε περιοχές και χώρες αντίστοιχα, και συνδέονται μεταξύ τους με ξένο κλειδί αλλά όχι με τον πίνακα fact\_earthquake. Ο πίνακας fact\_earthquakes συνδέεται με την διάσταση dim\_cities λαμβάνοντας υπ' όψη το πεδίο city\_id της κοντινότερης πόλης στο επίκεντρο του σεισμού. Οι διαστάσεις dim\_dates και dim\_times προσδιορίζουν την χρονική στιγμή του σεισμού, η διάσταση dim\_magnitude\_types προσδιορίζει τον τύπο μεγέθους του σεισμού. Η διάσταση dim\_subductions προσδιορίζει την ζώνη καταβύθισης στην περίπτωση που το επίκεντρο ενός σεισμού βρίσκεται οριζοντιογραφικά μέσα σε μία ζώνη καταβύθισης. Τέλος η διάσταση dim\_faults προσδιορίζει το πλήθος των σεισμικών ρηγμάτων στα οποία μπορεί να βρίσκεται οριζοντιογραφικά μέσα ένας σεισμός.



Σχήμα 5.1 Απεικόνιση σχήματος αποθήκης δεδομένων σεισμών

Τα ονόματα των πινάκων στον ενικό και χωρισμένα με κάτω παύλα ( ) και με πρόθεμα fact για των πίνακα των σεισμών και dim για τους πίνακες των διαστάσεων. Το πρόθεμα χρησιμοποιήθηκε για λόγους οργάνωσης των πινάκων.

Στην συνέχεια θα παρουσιαστούν οι πίνακες (tables) τα εναύσματα (triggers) και οι λειτουργίες (functions) που έχουν αναπτυχθεί την αποθήκη δεδομένων.

### 5.3.1 Date Dimension (dim\_dates)

Ο πίνακας dim\_dates (βλέπε Πίνακας 5.3) περιέχει πεδία για τον προσδιορισμό της ημερομηνίας. Τα δεδομένα ξεκινούν από την 1<sup>η</sup> Οκτωβρίου 2004, επομένως στον πίνακα αυτόν περάστηκαν ημερομηνίες από 1<sup>η</sup> Ιανουαρίου 2000 ως 31<sup>η</sup> Δεκεμβρίου 2100, έτσι ώστε να καλύπτουν τα δεδομένα για πολλά χρόνια ακόμα. Τα πεδία του πίνακα , με τους τύπους, τα μεγέθη και τις ιδιότητες τους περιγράφονται παρακάτω:

Όνομα πεδίου	Τύπος δεδομένων	Περιγραφή
date_id	character varying	Ημερομηνία (YYYYMMDD)
year	smallint	Χρονιά (2000-2100)
month	smallint	Μήνας (1-12)
day	smallint	Ημέρα (1-31)
season	character varying	Εποχή

Πίνακας 5.3 Περιγραφή πίνακα dim\_dates

Το πεδίο date\_id είναι τύπου character varying με μέγεθος 8 με σκοπό να κρατούν την ημερομηνία σε μορφή YYYYMMDD. Το πεδίο year περιέχει τον χρόνο (2000-2100), το πεδίο month περιέχει τον μήνα (1-12) και το πεδίο day περιέχει την μέρα (1-31) σε θετικούς ακέραιους αριθμούς. Το πεδίο season περιέχει την εποχή και δέχεται μία από τις τέσσερις λέξεις: Winter, Spring, Summer, Autumn.

### 5.3.2 Time Dimension (dim\_times)

Η Διάσταση των ωρών (βλέπε Πίνακας 5.4), περιέχει πεδία για τον προσδιορισμό της ώρας ενός σεισμού μέσα στην ημέρα. Τα πεδία της διάστασης των ωρών παρουσιάζονται στον

παρακάτω πίνακα. Τα πεδία hour, minute, second είναι τύπου smallint για εξοικονόμηση χώρου στην βάση δεδομένων.

Όνομα πεδίου	Τύπος δεδομένων	Ιδιότητα
time_id	character varying	Ωρα σε μορφή HHMMSS
hour	smallint	Ωρα (0-23)
minute	smallint	Λεπτά (0-59)
second	smallint	δευτερόλεπτα (0-59)

**Πίνακας 5.4 Περιγραφή πίνακα dim\_times**

Το πεδίο time\_id είναι τύπου character varying με μέγεθος 6 χαρακτήρων, με σκοπό να κρατάει την ώρα σε μορφή HHMMSS χωρίς να κόβονται τυχόν μηδενικά στην αρχή των χαρακτήρων. Τα πεδία hour (0-23), minute (0-59) και second (0-59) είναι τύπου smallint.

### 5.3.3 Crustal Fault Dimension (dim\_faults)

Μία εξίσου χρήσιμη πληροφορία για την αποθήκη δεδομένων είναι τα σεισμικά ρήγματα (βλέπε Πίνακας 5.5) στα οποία ανήκει μία σεισμική δραστηριότητα. Η πλήρης περιγραφική πληροφορία της διάστασης dim\_faults παρουσιάζεται στον παρακάτω πίνακα.

Όνομα πεδίου	Τύπος δεδομένων	Περιγραφή
fault_id	character varying	Μοναδικό κλειδί κάθε ρήγματος
fault_geometry	geometry(Polygon 4326)	Πολύγωνο ρήγματος
strike_min	double precision	Ελάχιστη οριζόντια γωνία
strike_avg	double precision	Μέση οριζόντια γωνία
strike_max	double precision	Μέγιστη οριζόντια γωνία
dip_min	double precision	Ελάχιστη κατακόρυφη γωνία κλίσης
dip_avg	double precision	Μέση κατακόρυφη γωνία κλίσης
dip_max	double precision	Μέγιστη κατακόρυφη γωνία κλίσης

rake_min	double precision	Ελάχιστη γωνία ολίσθησης
rake_avg	double precision	Μέση γωνία ολίσθησης
rake_max	double precision	Μέγιστη γωνία ολίσθησης
depth_min	double precision	Ελάχιστο βάθος
depth_max	double precision	Μέγιστο Βάθος
fault_type	character varying	Τύπος ρήγματος
fault_slip	character varying	Ολίσθηση ρήγματος

**Πίνακας 5.5 Περιγραφή πίνακα dim\_faults**

Το πεδίο fault\_id είναι το πρωτεύον κλειδί του πίνακα και το πεδίο fault\_geometry περιέχει την γεωμετρία του κάθε σεισμικού ρήγματος. Η γεωμετρία των σεισμικών ρηγμάτων είναι τύπου πολυγώνου στο σύστημα αναφοράς WGS '84 (EPSG:4326). Τα πεδία strike\_min, strike\_avg και strike\_max μας δείχνουν την ελάχιστη, μέση και μέγιστη (οριζόντια) γωνία διεύθυνσης ενός ρήγματος σε δεκαδικές μοίρες. Τα πεδία dip\_min, dip\_avg και dip\_max μας δείχνουν την ελάχιστη, μέση και μέγιστη γωνία (κατακόρυφη) κλίσης σε δεκαδικές μοίρες. Τα πεδία rake\_min, rake\_avg και rake\_max μας δείχνουν την ελάχιστη, μέση και μέγιστη γωνία ολίσθησης του ρήγματος σε δεκαδικές μοίρες. Τα πεδία depth\_min και depth\_max μας δείχνουν το ελάχιστο και μέγιστο βάθος το σεισμικού ρήγματος σε μέτρα. Το πεδίο fault\_type παίρνει τις τιμές LL, N, RL, R και αναφέρεται στον τύπο ρήγματος. Οι τύποι ρήγματος είναι κανονικό (N), ρήγμα επώθησης ή ανάστροφο ρήγμα (R), ρήγμα οριζόντιας μετατόπισης, δεξιόστροφο ή αριστερόστροφο (RL, LL) Το πεδίο fault\_slip παίρνει τις τιμές SS για strike-slip ή DS για dip-slip και δηλώνει την ολίσθηση του ρήγματος, όπως αυτοί αναφέρονται και στο άρθρο «What is a fault and what are the different types?» του Γεωλογικού Ινστιτούτου Ηνωμένων Πολιτειών (USGS).

#### **5.3.4 Subduction Discretization Dimension (dim\_subductions)**

Η διάσταση dim\_subductions (βλέπε Πίνακας 5..6) αναφέρεται στις ζώνες καταβυθίσεων των τεκτονικών πλακών και ο ρόλος της στην αποθήκη δεδομένων είναι για να ελέγξουμε αν ένας σεισμός ανήκει σε κάποια καταβύθιση. Η δομή της διάστασης dim\_subductions περιγράφεται στον παρακάτω πίνακα:

Όνομα πεδίου	Τύπος δεδομένων	Περιγραφή
subduction_id	character varying	Μοναδικό κλειδί ζώνης καταβύθισης
subduction_geometry	geometry(Polygon, 4326)	Γεωμετρία ζώνης καταβύθισης
slab_name	character varying	Ονομασία τεκτονικής πλάκας
depth_high	double precision	Μέγιστο βάθος ζώνης καταβύθισης
depth_low	double precision	Ελάχιστο βάθος ζώνης καταβύθισης

**Πίνακας 5.6 Περιγραφή πίνακα dim\_subductions**

Το πεδίο subduction\_id είναι το πρωτεύον κλειδί του πίνακα και το πεδίο subduction\_geometry είναι η γεωμετρία της καταβύθισης. Η γεωμετρία είναι τύπου πολυγώνου στο σύστημα αναφοράς WGS '84 (EPSG:4326). Το πεδίο slab\_name δηλώνει την τεκτονική πλάκα στην οποία ανήκουν οι ζώνες καταβυθίσεων. Τέλος τα πεδία depth\_high και depth\_low δηλώνουν το μέγιστο και το ελάχιστο βάθος του κάθε πολυγώνου των καταβυθίσεων.

### 5.3.5 Magnitude Type Dimension (dim\_magnitude\_types)

Η διάσταση dim\_magnitude\_types (βλέπε Πίνακας 5.7) μας δίνει πληροφορία σχετικά με τον τύπο μεγέθους του σεισμού. Στον παρακάτω πίνακα φαίνονται τα πεδία του πίνακα:

Όνομα πεδίου	Τύπος δεδομένων	Περιγραφή
magnitude_type	character varying	Τύπος μεγέθους σεισμών
description	character varying	Περιγραφή τύπου

**Πίνακας 5.7 Περιγραφή πίνακα dim\_magnitude\_types**

Το πεδίο magnitude\_type αναφέρεται στην κωδική ονομασία του τύπου μεγέθους του σεισμού, και το πεδίο description στην περιγραφή του τύπου αυτού.

### 5.3.6 City Dimension (dim\_cities)

Ο σκοπός της διάστασης των πόλεων (βλέπε Πίνακας 5.8) στην αποθήκη δεδομένων είναι για να έχουμε την κοντινότερη πόλη ενός σεισμού καθώς και την απόσταση αυτής από το επίκεντρο του. Τα πεδία του πίνακα dim\_cities παρουσιάζονται παρακάτω:

Όνομα πεδίου	Τύπος δεδομένων	Περιγραφή
city_id	smallserial	Μοναδικό κλειδί κάθε πόλης
city_geometry	geometry(Point, 4326)	Γεωμετρία πόλης
city_name	character varying	Ονομασία πόλης
region_id	character varying	κλειδί περιοχής στην οποία ανήκει κάθε πόλη (FK)

**Πίνακας 5.8 Περιγραφή πίνακα dim\_cities**

Το πεδίο city\_id είναι smallint για εξοικονόμηση χώρου και πρωτεύων κλειδί του πίνακα dim\_cities. Το πεδίο city\_geometry είναι η γεωμετρία κάθε πόλης ως σημείο στο σύστημα αναφοράς WGS '84 (EPSG:4326). Το πεδίο city\_name περιέχει την ονομασία της πόλης και το πεδίο region\_id είναι ξένο κλειδί στον πίνακα dim\_regions και αναφέρεται στην περιοχή στην οποία ανήκει μία πόλη.

### 5.3.7 Region Dimension (dim\_regions)

Η διάσταση dim\_regions (βλέπε Πίνακας 5.9) αναφέρεται στις ευρύτερες περιοχές των πόλεων και τα πεδία της παρουσιάζονται στον παρακάτω πίνακα:

Όνομα πεδίου	Τύπος δεδομένων	Περιγραφή
region_id	character varying	Μοναδικό κλειδί κάθε περιοχής
region_geometry	geometry(MultiPolygon, 4326)	Γεωμετρία κάθε περιοχής
country_id	character varying	Κλειδί χώρας στην οποία ανήκει κάθε περιοχή (FK)
region_name	character varying	Ονομασία περιοχής

**Πίνακας 5.9 Περιγραφή πίνακα dim\_regions**



Το πεδίο `region_id` είναι πρωτεύων κλειδί του πίνακα `dim_cities`. Το πεδίο `region_geometry` είναι η γεωμετρία της κάθε περιοχής ως `multi-polygon` στο σύστημα αναφοράς WGS '84 (EPSG:4326). Το πεδίο `region_name` περιέχει την ονομασία της περιοχής και το πεδίο `country_id` είναι ξένο κλειδί στον πίνακα `dim_countries` και αναφέρεται στην χώρα στην οποία ανήκει μία περιοχή.

### 5.3.8 Country Dimension (`dim_countries`)

Ο σκοπός της διάστασης των χωρών (βλέπε Πίνακας 5.10) στην αποθήκη δεδομένων είναι για να έχουμε την χώρα στην οποία έγινε ένας σεισμός. Τα πεδία του πίνακα `dim_countries` παρουσιάζονται παρακάτω:

Όνομα πεδίου	Τύπος δεδομένων	Περιγραφή
<code>country_id</code>	<code>character varying</code>	Μοναδικό κλειδί κάθε χώρας
<code>country_geometry</code>	<code>geometry(MultiPolygon, 4326)</code>	Γεωμετρία κάθε χώρας
<code>country_name</code>	<code>character varying</code>	Ονομασία κάθε χώρας

**Πίνακας 5.10** Περιγραφή πίνακα `dim_countries`

Το πεδίο `country_id` είναι πρωτεύων κλειδί του πίνακα `dim_countries`. Το πεδίο `country_geometry` είναι η γεωμετρία κάθε χώρας ως `multi-polygon` στο σύστημα αναφοράς WGS '84 (EPSG:4326). Το πεδίο `country_name` περιέχει την ονομασία της κάθε χώρας.

### 5.3.9 Fact Earthquake (`fact_earthquakes`)

Ο πίνακας `fact_earthquakes` (βλέπε Πίνακας 5.11) είναι ο πίνακας των γεγονότων, ο οποίος φιλοξενεί τους σεισμούς με τα χαρακτηριστικά τους και τα ξένα κλειδιά των διαστάσεων του σχήματος. Αναλυτικά η σύνθεση του `fact_earthquakes` φαίνεται στην παρακάτω πίνακα:

Όνομα πεδίου	Τύπος δεδομένων	Περιγραφή
<code>earthquake_id</code>	<code>bigserial</code>	Μοναδικό κλειδί κάθε σεισμού
<code>date_id</code>	<code>character varying</code>	Ημερομηνία σεισμού (FK)



time_id	character varying	Ωρα σεισμού (FK)
subduction_id	integer	Κλειδί ζώνης καταβύθισης (FK)
city_id	integer	Κλειδί κοντινότερης πόλης (FK)
city_distance	double precision	Απόσταση σε χλμ κοντινότερης πόλης
magnitude_type	smallint	Τύπος μεγέθους σεισμού (FK)
epicenter	geometry(Point, 4326)	Επίκεντρο σεισμού
magnitude	double precision	Μέγεθος σεισμού σε κλίμακα Richter
hypocenter_depth	double precision	Εστιακό βάθος σεισμού
fault_array	character varying[]	Λίστα σεισμικών ρηγμάτων (FK)

**Πίνακας 5.11 Περιγραφή πίνακα fact\_earthquakes**

Το πεδίο earthquake\_id είναι το πρωτεύον κλειδί του πίνακα και ως εκ τούτου μοναδικό για κάθε σεισμό ο οποίος εισάγεται μέσα στον πίνακα. Το πεδίο date\_id είναι ξένο κλειδί και αναφέρεται στην διάσταση των ημερομηνιών dim\_dates και προσδιορίζει την ημερομηνία την οποία έλαβε χώρα ο σεισμός. Το πεδίο dim\_time είναι ξένο κλειδί και αναφέρεται στην διάσταση του χρόνου dim\_times και προσδιορίζει ακριβή ώρα την οποία έλαβε χώρα ο σεισμός μέσα στην ημέρα. Το πεδίο subduction\_id είναι ξένο κλειδί, αναφέρεται στην διάσταση των καταβυθίσεων dim\_subductions και προσδιορίζει το αν κάποιος σεισμός βρίσκεται σε σημείο καταβύθισης τεκτονικής πλάκας.

Το πεδίο city\_id είναι ξένο κλειδί και αναφέρεται στην διάσταση των πόλεων dim\_city και προσδιορίζει την κοντινότερη πόλη στην οποία έλαβε χώρα ο σεισμός. Η εύρεση της κοντινότερης πόλης και η απόσταση της από το επίκεντρο του σεισμού, γίνεται αυτόματα από την αποθήκη δεδομένων με την λειτουργία fn\_nearest\_city.

Το πεδίο city\_distance προσδιορίζει την απόσταση σε χιλιόμετρα του επίκεντρου του σεισμού από την κοντινότερη πόλη. Το πεδίο magnitude\_type είναι ξένο κλειδί, αναφέρεται στην διάσταση των τύπων μεγέθους dim\_magnitude\_types, και προσδιορίζει τον τύπο μεγέθους του σεισμού. Το πεδίο epicenter αναφέρεται στο επίκεντρο του σεισμού και είναι γεωμετρίας τύπου σημείου στο παγκόσμιο σύστημα συντεταγμένων WGS '84

(EPSG:4326). Το πεδίο magnitude προσδιορίζει το μέγεθος του σεισμού σε μονάδες της κλίμακας Richter. Το πεδίο hypocenter\_depth προσδιορίζει το εστιακό βάθος του σεισμού.

Το πεδίο fault\_array είναι ξένο κλειδί και αναφέρεται στην διάσταση dim\_faults και προσδιορίζει το ρήγμα ή τα ρήγματα στα οποία ανήκει ο σεισμός. Ένας σεισμός ανήκει σε ένα ρήγμα αν οι συντεταγμένες του σημείου του βρίσκονται χωρικά εντός του πολυγώνου του ρήγματος. Επειδή όμως χωρικά τυγχάνει ένα σημείο να βρίσκεται ταυτόχρονα σε περισσότερα του ενός πολύγωνα (ρηγμάτων) τα οποία επικαλύπτονται, το πεδίο fault\_array είναι μία λίστα που δέχεται περισσότερες από μία τιμές.

## 5.4 Συναρτήσεις και Εναύσματα (functions & triggers)

Στο κεφάλαιο αυτό περιγράφονται οι συναρτήσεις και τα εναύσματα τα οποία αναπτύχθηκαν. Οι συναρτήσεις και τα εναύσματα δημιουργήθηκαν για να αυτοματοποιηθούν κάποιες λειτουργίες κατά την εισαγωγή των δεδομένων, αυτοματοποιώντας έτσι τον υπολογισμό δεδομένων τα οποία δεν υπάρχουν και πρέπει να υπολογιστούν, διασφαλίζοντας έτσι την ποιότητα τους και αποφεύγοντας την περίπτωση ένας χρήστης να εισάγει λάθος δεδομένα. Αυτές οι λειτουργίες είναι η εύρεση της περιοχής μίας πόλης, η εύρεση της κοντινότερης πόλης σε έναν σεισμό και ο υπολογισμός της απόστασης αυτής σε χιλιόμετρα, η εύρεση σεισμικών ρηγμάτων ενός σεισμού και η εύρεση ζωνών καταβυθίσεων ενός σεισμού.

### 5.4.1 Λειτουργία σύνδεσης πόλεων με περιοχές

Έχοντας φορτωμένα τα δεδομένα των περιοχών και των πόλεων, επόμενο βήμα είναι η συμπλήρωση του πεδίου region\_id της διάστασης dim\_cities κατάλληλα. Το πεδίο θα πρέπει να περιέχει το id της περιοχής στην οποία ανήκει μία πόλη και να συμπληρώνεται αυτόματα από την αποθήκη δεδομένων. Για την εύρεση του id της περιοχής αναπτύχθηκε η λειτουργία fn\_cities\_region\_id στην οποία, γίνεται χρήση της συνάρτησης ST\_WITHIN του PostGIS, όπου από τον πίνακα dim\_regions επιλέγεται το region\_id της εγγραφής αυτής στην οποία βρίσκεται μέσα ένα σημείο πόλης. Επίσης για να διασφαλίσουμε ότι το ερώτημα

θα επιστρέψει ένα μόνο `region_id` περιορίζουμε τον αριθμό αποτελεσμάτων του ερωτήματος σε 1 με την χρήση της εντολής `LIMIT 1`.

```
SELECT region_id
FROM dim_regions
WHERE st_within(NEW.city_geometry, region_geometry)
LIMIT 1
```

#### **Κώδικας 5 Ερώτημα sql επιστροφής id περιοχής**

Η παραπάνω λειτουργία εκτελείται πριν από την εισαγωγή μίας εγγραφής στον πίνακα `dim_cities`, ή πριν την ενημέρωση του πεδίου `city_geometry` κάποιας εγγραφής του πίνακα `dim_cities`.

### **5.4.2 Λειτουργία σύνδεσης σεισμών με ζώνες καταβύθισεων**

Επόμενο βήμα ήταν η συμπλήρωση του πεδίου `subduction_id` του πίνακα `fact_earthquakes`, με το `id` της ζώνης καταβύθισης μέσα στην οποία βρίσκεται το επίκεντρο ενός σεισμού. Ομοίως με την λειτουργία `fn_cities_region_id`, αναπτύχθηκε τη λειτουργία `fn_earthquakes_subduction_id` η οποία επιστρέφει το `id` της ζώνης καταβύθισης με το ερώτημα που φαίνεται παρακάτω:

```
SELECT subduction_id
FROM dim_subductions
WHERE st_within(NEW.epicenter, subduction_geometry)
LIMIT 1
```

#### **Κώδικας 6 Ερώτημα sql επιστροφής id ζώνης καταβύθισης**

Η παραπάνω λειτουργία εκτελείται πριν από την εισαγωγή μίας εγγραφής στον πίνακα `fact_earthquakes`, ή πριν την ενημέρωση του πεδίου `epicenter` κάποιας εγγραφής του πίνακα `fact_earthquakes`.

### **5.4.3 Λειτουργία εύρεσης μικρότερης απόστασης μεταξύ επίκεντρου σεισμού και πόλης**

Στην συνέχεια σειρά έχει η εύρεση της κοντινότερης πόλης από το επίκεντρο ενός σεισμού και η απόσταση αυτή σε χιλιόμετρα, έτσι ώστε να συμπληρώνονται τα πεδία `nearest_city_id` και `city_distance` του πίνακα `fact_earthquakes`. Την δουλειά αυτή την κάνει η λειτουργία `fn_nearest_city`, η οποία αρχικά από το επίκεντρο του σεισμού θα επιλέξει όλα τα σημεία των πόλεων τα οποία βρίσκονται εντός ακτίνας 26.1 μοιρών με την χρήση της συνάρτησης `st_expand`, και από αυτά θα επιλέξει την πόλη με την μικρότερη απόσταση με την συνάρτηση `st_distancesphere`, και θα επιστρέψει το `id` και την απόσταση αυτή. Η τιμή των

26.1 μοιρών επιλέχθηκε καθώς κάλυπτε την μεγαλύτερη απόσταση επίκεντρου σεισμού από πόλη. Επίσης με τον τρόπο αυτό αποφεύγουμε να ελέγξουμε το επίκεντρο του σεισμού με κάθε πόλη, πράγμα που θα είχε πολυπλοκότητα  $O(n^2)$ , και θα καθυστερούσε την εισαγωγή και επεξεργασία των εγγραφών στον πίνακα `fact_earthquakes`.

```
WITH filtered_cities AS
(
SELECT city_id, city_geometry
FROM dim_cities
WHERE city_geometry && st_expand(new.epicenter, 26.1)
)
SELECT city_id, st_distancesphere(new.epicenter, city_geometry) / 1000
AS distance
INTO new.nearest_city_id, new.city_distance
FROM filtered_cities
ORDER BY distance
LIMIT 1;
```

**Κώδικας 7 Ερώτημα sql επιστροφής id και απόστασης κοντινότερης πόλης**

Η παραπάνω λειτουργία εκτελείται πριν από την εισαγωγή μίας εγγραφής στον πίνακα `fact_earthquakes`, ή πριν την ενημέρωση του πεδίου `epicenter` κάποιας εγγραφής του πίνακα `fact_earthquakes`.

#### 5.4.4 Λειτουργία σύνδεσης σεισμών με σεισμικά ρήγματα

Η λειτουργία σύνδεσης των σεισμών με τα σεισμικά ρήγματα, έχει σαν σκοπό να συμπληρώσει το πεδίο `fault_array` με το `id` ή τα `id` των σεισμικών ρηγμάτων, στα οποία μπορεί να βρίσκεται οριζοντιογραφικά ένας σεισμός. Η λειτουργία αυτή επιστρέφει όλα τα δυνατά `id` των ρηγμάτων σε μορφή λίστας και τα εισάγει στο πεδίο `fault_array`. Η εύρεση των `id` των ρηγμάτων γίνεται με την συνάρτηση `st_contains`, η οποία βρίσκει όλα τα ρήγματα στα οποία βρίσκεται ένας σεισμός, και τα κάνει λίστα με την συνάρτηση `array_agg`.

```
SELECT array_agg(fault_id)
INTO new.fault_array
FROM dim_faults
WHERE st_contains(fault_geometry, new.epicenter);
```

**Κώδικας 8 Ερώτημα sql εύρεσης λίστας ρηγμάτων**

Η παραπάνω λειτουργία εκτελείται πριν από την εισαγωγή μίας εγγραφής στον πίνακα `fact_earthquakes`, ή πριν την ενημέρωση του πεδίου `epicenter` κάποιας εγγραφής του πίνακα `fact_earthquakes`.

## 5.5 Ερωτήματα SQL

Η αποθήκη δεδομένων σχεδιάστηκε για να καλύπτει τόσο απλά όσο και πιο πολύπλοκα ερωτήματα sql συναρτήσει του χώρου αλλά και του χρόνου. Για την αξιοποίηση των παραπάνω δεδομένων και όλου του φάσματος των δυνατοτήτων της αποθήκης δεδομένων επιλέχθηκαν τα παρακάτω ερωτήματα:

1. Σεισμοί οι οποίοι βρίσκονται μέσα σε ρήγμα
2. Σεισμοί οι οποίοι βρίσκονται μέσα σε ζώνη καταβύθισης
3. 10 μεγαλύτεροι σεισμοί κοντά στην Πάτρα την τελευταία δεκαετία
4. Ποσοστό σεισμών ανά χώρα
5. Ποσοστό ρηγμάτων ανά χώρα
6. 20 πιο σεισμογενείς πόλεις
7. 20 μεγαλύτεροι σεισμοί
8. Ποσοστό σεισμών μέσα σε ρήγμα και ποσοστό σεισμών μέσα σε ζώνη καταβύθισης
9. Συνολικοί και μεγαλύτεροι σεισμοί ανά έτος
10. Συνολικοί και μεγαλύτεροι σεισμοί ανά μήνα
11. Συνολικοί και μεγαλύτεροι σεισμοί ανά εποχή

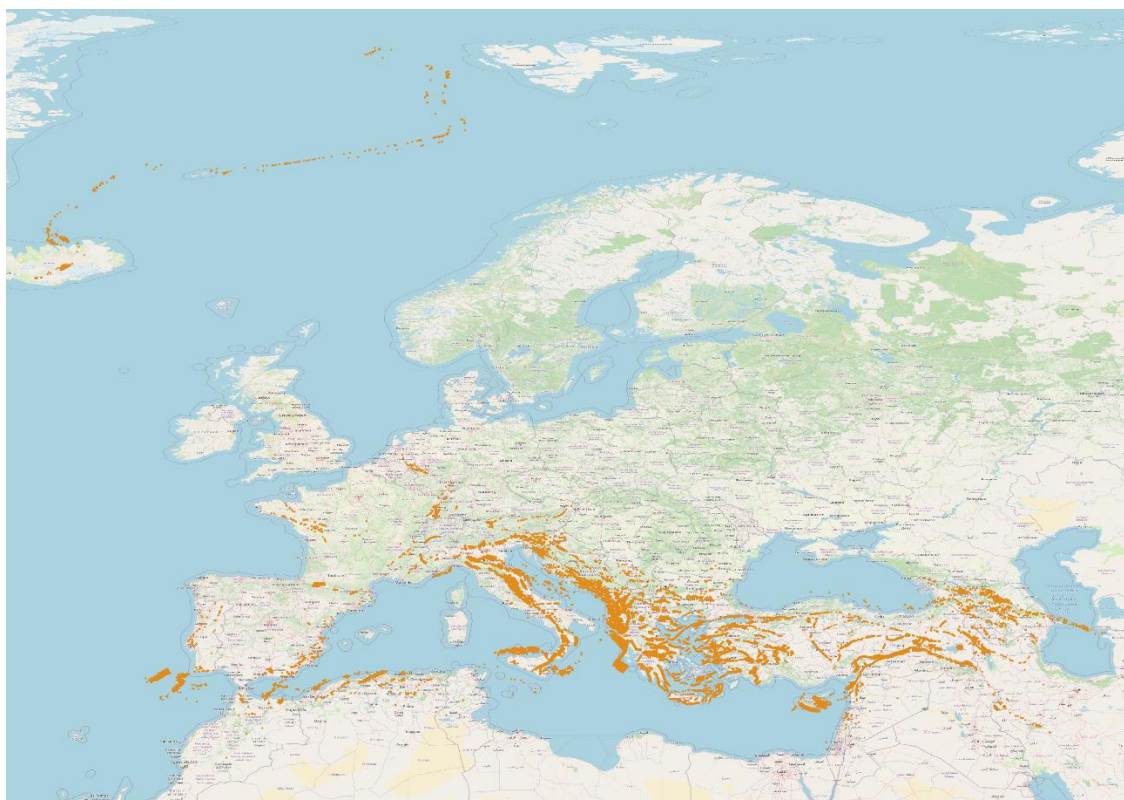
Σημειώνεται ότι για τα ερωτήματα 10 και 11 δεν λήφθηκαν υπ' όψη οι σεισμοί των ετών 2004 και 2023 διότι τα δεδομένα τους δεν καλύπτουν ολόκληρο το έτος, επομένως θα είχαμε λανθασμένο αποτέλεσμα. Τα ερωτήματα 1, 2, 3, 4, 5, 6 και 8 θα παρουσιαστούν σε θεματικούς χάρτες, ενώ τα ερωτήματα 7, 9, 10 και 11 σε διαγράμματα.

## 6. Αποτελέσματα Ερωτημάτων SQL

Σε αυτό το κεφάλαιο θα παρουσιαστούν τα αποτελέσματα των ερωτημάτων sql. Τα ερωτήματα 1 -7 εξήχθησαν σαν GeoJSON από την αποθήκη δεδομένων και στην συνέχεια με την χρήση του λογισμικού QGIS μετατράπηκαν σε θεματικούς χάρτες όπως αυτοί φαίνονται παρακάτω.

Σημειώνεται ότι για τα ερωτήματα 10 και 11 δεν λήφθηκαν υπ' όψη οι σεισμοί των ετών 2004 και 2023 διότι τα δεδομένα τους δεν καλύπτουν ολόκληρο το έτος, επομένως θα είχαμε λανθασμένο αποτέλεσμα.

Το αποτέλεσμα του ερωτήματος 1 (βλέπε Εικόνα 6.1) μας απεικονίζει τους σεισμούς με πορτοκαλί χρώμα, οι οποίοι βρίσκονται μέσα σε κάποιο ρήγμα. Σαν υπόβαθρο χάρτη χρησιμοποιήθηκε ο χάρτης του Open Street Maps.



Εικόνα 6.1 Αποτέλεσμα ερωτήματος 1 - σεισμοί εντός ρηγμάτων

Το αποτέλεσμα του ερωτήματος 2 (βλέπε Εικόνα 6.2) μας απεικονίζει τους σεισμούς με ανοιχτό πράσινο χρώμα, οι οποίοι βρίσκονται μέσα σε κάποια ζώνη καταβύθισης, σαν

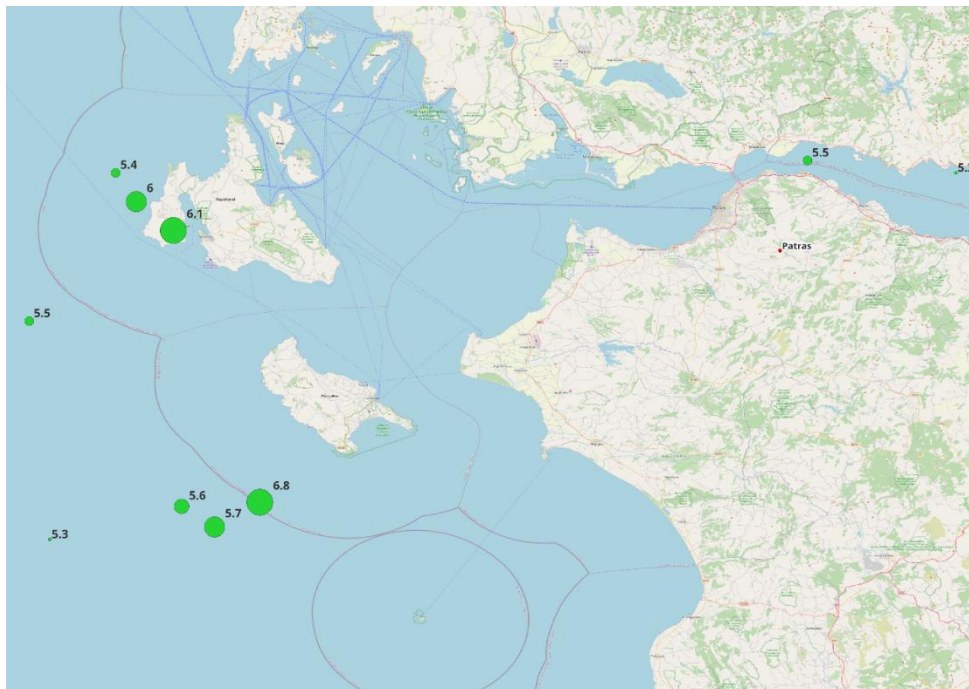


υπόβαθρο χάρτη χρησιμοποιείται ο χάρτης δορυφορικών εικόνων της Google (Google Satellite).



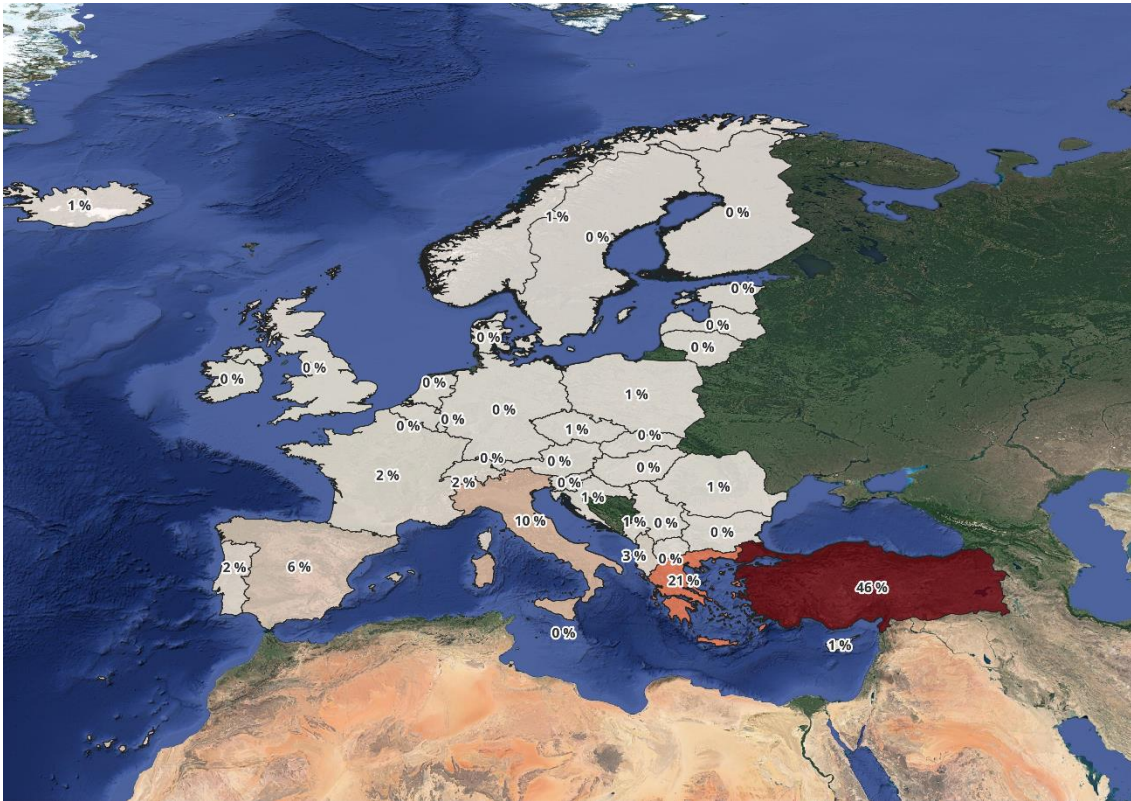
Εικόνα 6.2 Αποτέλεσμα ερωτήματος 2 - σεισμοί εντός ζωνών καταβυθίσεων

Το αποτέλεσμα του ερωτήματος 3 (βλέπε Εικόνα 6.3) μας απεικονίζει τους μεγαλύτερους σεισμούς οι οποίοι έγιναν κοντά στην Πάτρα με πράσινο χρώμα. Σαν υπόβαθρο χάρτη χρησιμοποιήθηκε ο χάρτης του Open Street Maps.



Εικόνα 6.3 Αποτέλεσμα ερωτήματος 3 - 10 μεγαλύτεροι σεισμοί της τελευταίας δεκαετίας κοντά στην Πάτρα

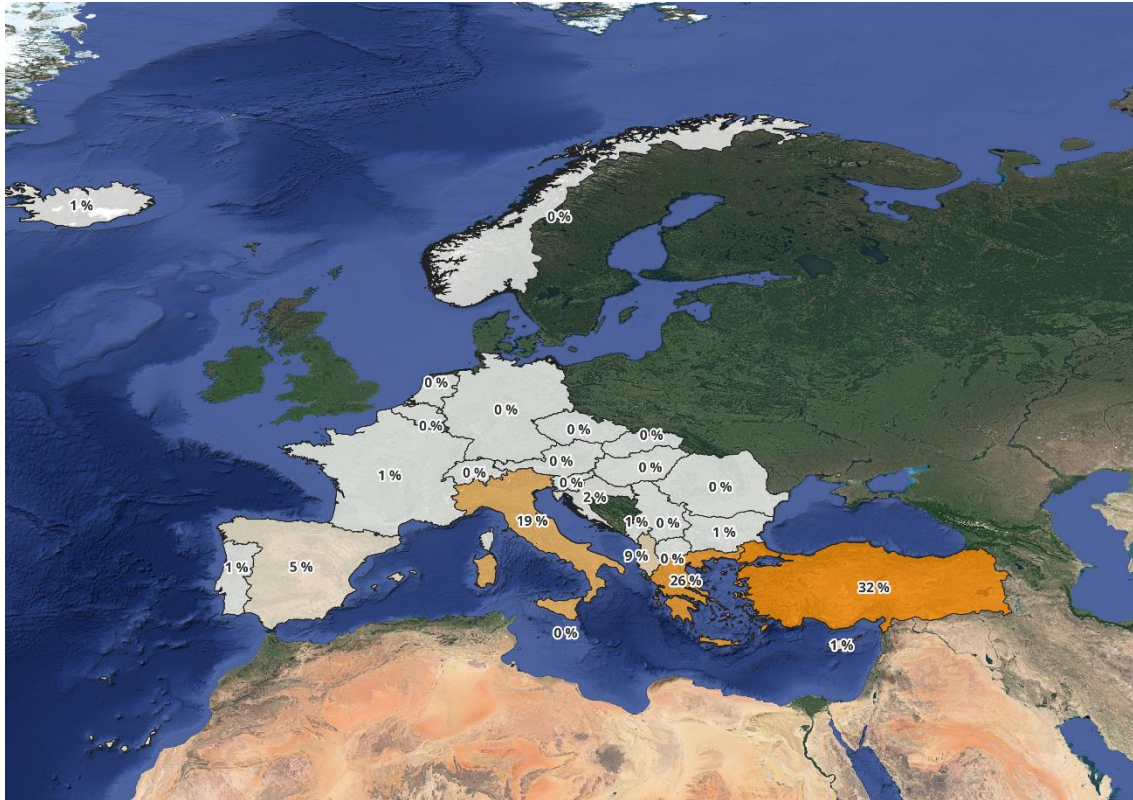
Το αποτέλεσμα του ερωτήματος 4 (βλέπε Εικόνα 6.4) μας απεικονίζει το ποσοστό επί τις εκατό των σεισμών που έχουν συμβεί σε κάθε χώρα, σαν υπόβαθρο χάρτη χρησιμοποιείται ο χάρτης δορυφορικών εικόνων της Google (Google Satellite). Το ποσοστό αναφέρεται σε σεισμούς οι οποίοι βρίσκονται εντός της ηπειρωτικής χώρας ή των νησιών μιας χώρας και δεν συμπεριλαμβάνει του σεισμούς οι οποίοι βρίσκονται εντός εθνικών υδάτων.



**Εικόνα 6.4 Αποτέλεσμα ερωτήματος 4 - ποσοστό σεισμών ανά χώρα**

Το αποτέλεσμα του ερωτήματος 5 (βλέπε Εικόνα 6.5) μας απεικονίζει το ποσοστό επί τις εκατό των ρηγμάτων που έχει κάθε χώρα, σαν υπόβαθρο χάρτη χρησιμοποιείται ο χάρτης δορυφορικών εικόνων της Google (Google Satellite).





Εικόνα 6.5 Αποτέλεσμα ερωτήματος 5 - ποσοστό ρηγμάτων ανά χώρα

Το αποτέλεσμα του ερωτήματος 6 (βλέπε Εικόνα 6.6) μας τις 20 πιο σεισμογενείς πόλεις της Ευρώπης, σαν υπόβαθρο χάρτη χρησιμοποιείται ο χάρτης δορυφορικών εικόνων της Google (Google Satellite).



Εικόνα 6.6 Αποτέλεσμα ερωτήματος 6 - 20 πιο σεισμογενείς πόλεις της Ευρώπης

Το αποτέλεσμα του ερωτήματος 7 (βλέπε Εικόνα 6.7) μας απεικονίζει τους 20 μεγαλύτερους σεισμούς που έχουν γίνει στην Ευρώπη, σαν υπόβαθρο χάρτη χρησιμοποιείται ο χάρτης δορυφορικών εικόνων της Google (Google Satellite).



Εικόνα 6.7 Αποτέλεσμα ερωτήματος 7 - 20 μεγαλύτεροι σεισμοί.

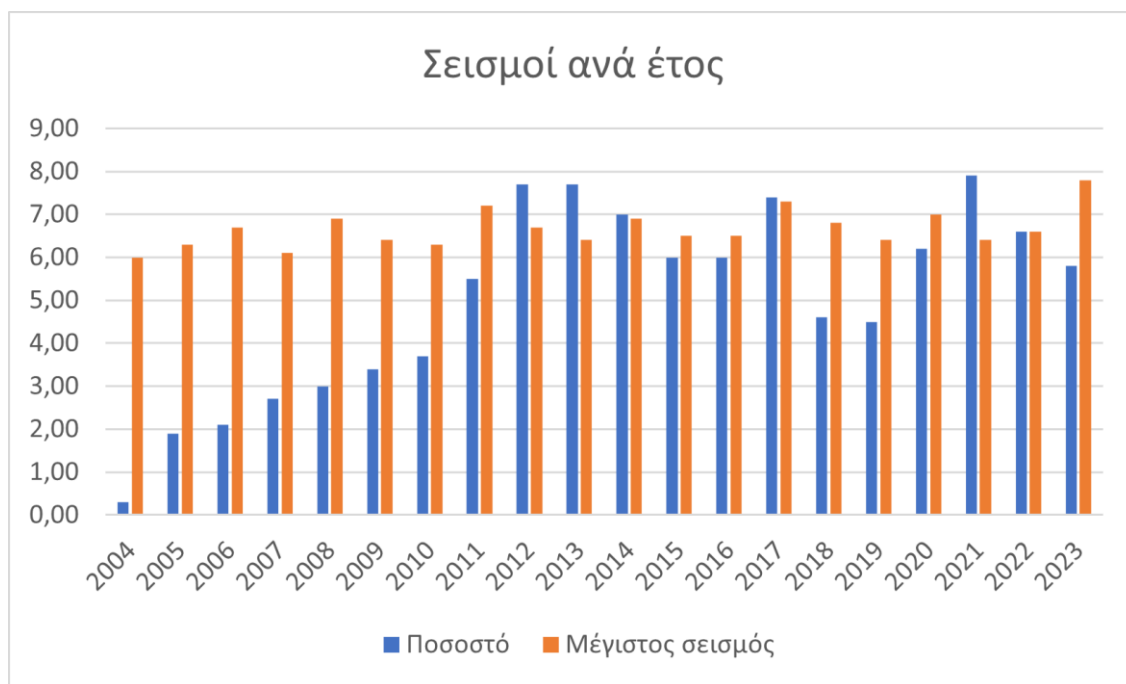
Το αποτέλεσμα του ερωτήματος 8 (βλέπε Πίνακας 6.1) μας δείχνει ότι το 10.75 % από όλους τους σεισμούς βρίσκεται μέσα σε κάποια ζώνη καταβύθισης, Ενώ το 19.85 % των σεισμών βρίσκεται μέσα σε κάποιο σεισμικό ρήγμα.

Παράμετρος	Τιμή
Σεισμοί σε ζώνη καταβύθισης	33024
Ποσοστό σεισμών σε ζώνη καταβύθισης	10.75 %
Σεισμοί σε ρήγμα	61001
Ποσοστό σεισμών σε ρήγμα	19.85 %

Πίνακας 6.1 Αποτέλεσμα ερωτήματος 8 – ποσοστό σεισμών σε ρήγμα & ποσοστό σεισμών σε ζώνη καταβύθισης

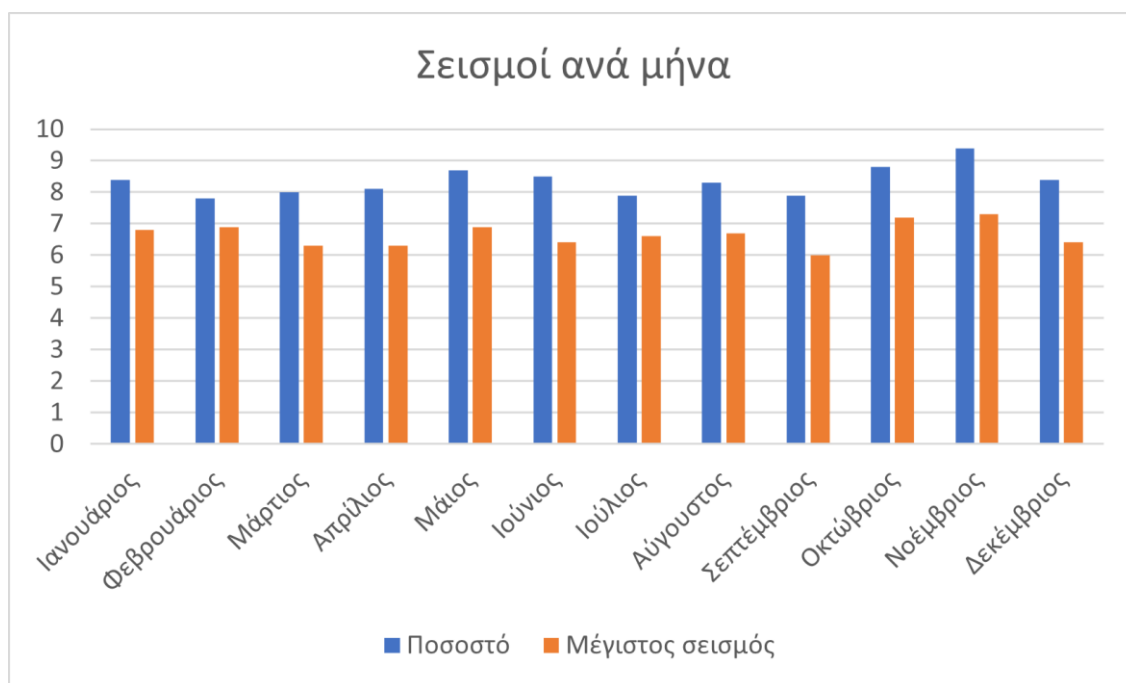
Το αποτέλεσμα του ερωτήματος 9 (βλέπε Εικόνα 6.8) μας δείχνει την συχνότητα των σεισμών και τον μέγιστο σεισμό ανά έτος.





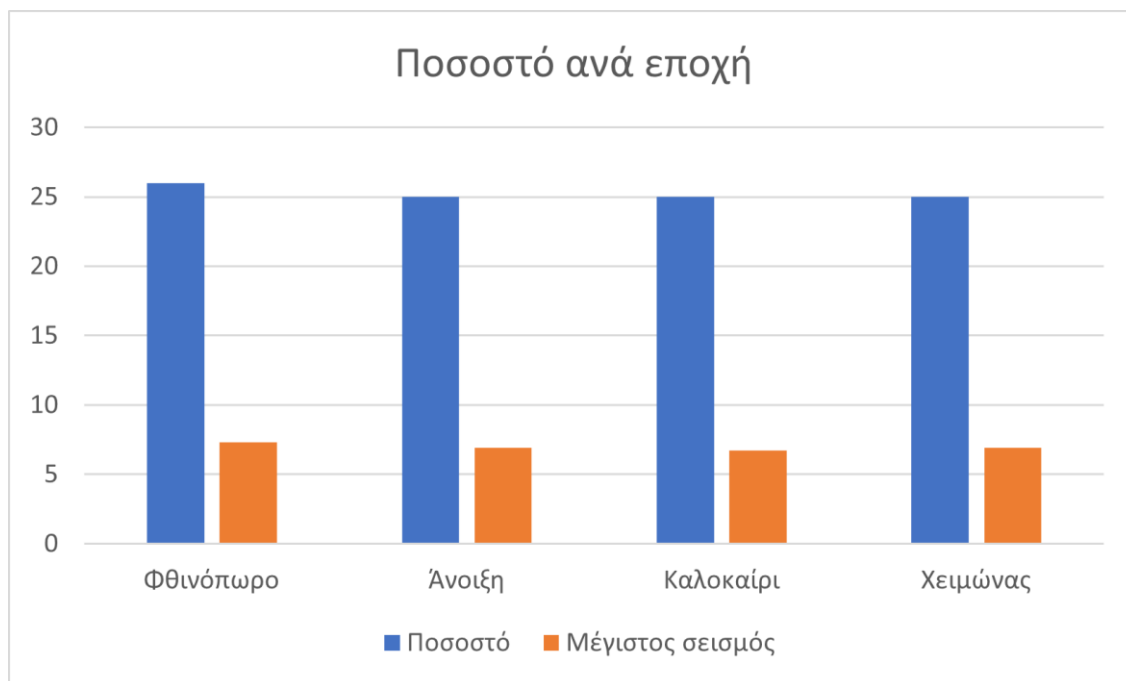
Εικόνα 6.8 Αποτέλεσμα ερωτήματος 9 - σεισμοί ανά έτος

Το αποτέλεσμα του ερωτήματος 10 (βλέπε Εικόνα 6.9) μας δείχνει την συχνότητα των σεισμών και τον μέγιστο σεισμό ανά μήνα.



Εικόνα 6.9 Αποτέλεσμα ερωτήματος 10 - σεισμοί ανά μήνα

Το αποτέλεσμα του ερωτήματος 9 (βλέπε Εικόνα 6.10) μας δείχνει την συχνότητα των σεισμών και τον μέγιστο σεισμό ανά εποχή.



Εικόνα 6.10 Αποτέλεσμα ερωτήματος 11 - σεισμοί ανά μήνα

## 7. Συμπεράσματα

Η αποθήκη δεδομένων που αναπτύχθηκε ανταποκρίνεται εύστοχα σε ζητήματα ανάλυσης και διαχείρισης σεισμολογικών δεδομένων, ενώ παράλληλα εγγυάται την ακεραιότητα σύνθετων δεδομένων σε διαφορετικές περιπτώσεις χρήσης. Η πολυδιάστατη προσέγγιση των γεωχωρικών δεδομένων επιτρέπει την ανάλυση των δεδομένων από διαφορετικές οπτικές γωνίες, και διευκολύνει την απεικόνιση τους σε θεματικούς χάρτες.

Η ανάπτυξη της αποθήκης δεδομένων σε PostgreSQL παρέχει την δυνατότητα εκτέλεσης γεωχωρικών ερωτημάτων, επιτρέπει την επεκτασιμότητα τόσο στην συμπερίληψη και άλλων ηπειρών σε ξεχωριστά ή και στο ίδιο σχήμα, όσο και την συμπερίληψη επιπρόσθετων δεδομένων ανεξαρτήτως τύπου δομής και μορφής.

Παρατηρούμε μία συσχέτιση μεταξύ των ερωτημάτων 4 και 5 μεταξύ Ιταλίας, Ελλάδας και Τουρκίας, δηλαδή των σεισμών και των ρηγμάτων ανά χώρα. Η συσχέτιση αυτή μπορεί να υποστηριχθεί ακόμα περισσότερο αν ληφθούν υπ' όψη και οι σεισμοί εντός εθνικών χωρικών υδάτων, όπου θα υπάρχουν περισσότεροι σεισμοί ποσοστιαία σε Ελλάδα και Ιταλία.

Οι χώρες της Μεσογείου είναι οι πιο σεισμογενείς της Ευρώπης, έχοντας το μεγαλύτερο ποσοστό σεισμών, σεισμικών ρηγμάτων και ζωνών καταβυθίσεων πράγμα που επαληθεύεται από τα δεδομένα. Ο χρόνος και η εποχή δεν επηρεάζουν την γένεση των σεισμών όπως προκύπτει από το αποτέλεσμα του ερωτήματος 10.

Η αποθήκη δεδομένων κάλυψε ευρύ εύρος ερωτημάτων και έδωσε πολλές διαφορετικές οπτικές γωνίες στην ανάλυση και διαχείριση των σεισμολογικών δεδομένων. Πολλά πράγματα όμως θα μπορούσαν να συνεχίσουν και να επεκτείνουν την ανάπτυξη της αποθήκης δεδομένων. Σαν δεδομένα ή διάσταση θα μπορούσαν να ληφθούν υπ' όψη και τα χωρικά ύδατα των χωρών, πράγμα που θα έδινε μεγαλύτερη ακρίβεια στα αποτελέσματα των ερωτημάτων. Επίσης θα μπορούσαν να εισαχθούν σεισμοί και δεδομένα από περισσότερες ηπείρους, και από περισσότερες πηγές δεδομένων με σκοπό την επέκταση της αποθήκης δεδομένων. Τέλος θα μπορούσαν να ληφθούν υπ' όψη δεδομένα για σεισμικές ακολουθίες και δεδομένα για ηφαίστεια, πράγμα που θα πολλαπλασίαζε τις αναλυτικές δυνατότητες της αποθήκης δεδομένων.



## Βιβλιογραφία

Ακολουθούν οι βιβλιογραφικές αναφορές (πηγές) της Εργασίας.

Δρούτσος Θ., (2000). Γεωλογία: Αρχές και Εφαρμογές. Αθήνα: Εκδόσεις Leader Books.

Έγκριση εφαρμογής και χρήσης Ευρωκωδίκων σε συνδυασμό με τα αντίστοιχα Εθνικά Προσαρτήματα. Εφημερίδα της Κυβέρνησης (ΦΕΚ 1457 Β/5-6-2014).

Ελληνικός Αντισεισμικός Κανονισμός 2000 (ΕΑΚ 2000). Εφημερίδα της Κυβέρνησης (ΦΕΚ 2184 Β/20-12-1999).

Νέος Ελληνικός Αντισεισμικός Κανονισμός (ΝΕΑΚ). Εφημερίδα της Κυβέρνησης (ΦΕΚ 613Β/12-10-1992).

Παπανικολάου Δ. & Σιδέρης Χ., (2009). Γεωλογία η Επιστήμη της Γης. Αθήνα: Εκδόσεις Πατάκη.

Περί αντισεισμικού κανονισμού οικοδομικών έργων. Εφημερίδα της Κυβέρνησης (ΦΕΚ 36Α/26-2-1959).

Τροποποίηση και συμπλήρωση του ΒΔ της 19/26 Φεβρουαρίου 1959 «περί αντισεισμικού κανονισμού οικοδομικών έργων». Εφημερίδα της Κυβέρνησης (ΦΕΚ 239Β/16-4-1984).

Basili R., Danciu L., Beauval C., Sesetyan K., Vilanova S., Adamia S., Arroucau P., Atanackov J., Baize S., Canora C., Caputo R., Carafa M., Cushing M., Custódio S., Demircioglu Tumsa M., Duarte J., Ganas A., García-Mayordomo J., Gómez de la Peña L., Gràcia E., Jamšek Rupnik P., Jomard H., Kastelic V., Maesano F., Martín-Banda R., Martínez-Loriente S., Neres M., Perea H., Sket-Motnikar B., Tiberti M., Tsereteli N., Tsironi V., Vallone R., Vanneste K., Zupančič P. (2022). European Fault-Source Model 2020 (EFSM20). Istituto Nazionale di Geofisica e Vulcanologia (INGV). <https://doi.org/10.13127/efsm20>

Bormann, P. (2011). Earthquake, Magnitude. In: Gupta, H.K. (eds) Encyclopedia of Solid Earth Geophysics. Encyclopedia of Earth Sciences Series. Springer, Dordrecht. [https://doi.org/10.1007/978-90-481-8702-7\\_3](https://doi.org/10.1007/978-90-481-8702-7_3)

Bormann P., Wendt S., Di Giacomo D. (2013). Seismic Sources and Source Parameters Chapter 3 (Version August 2013; DOI: 10.2312/GFZ.NMSOP-2\_ch3).

- Countries 2020 (2023, Μάιος 5) Ανακτήθηκε από <https://ec.europa.eu/eurostat/web/gisco/geodata/reference-data/administrative-units-statistical-units/countries#countries20>
- Earle S., (2015). Physical Geology. Victoria, B.C.: BCcampus. Ανακτήθηκε από <https://opentextbc.ca/geology>
- EFSM20 (2023, Φεβρουάριος 10). Ανακτήθηκε από <https://seismofaults.eu/efsm20data>
- Han, J.; Kamber, M. & Pei, J. (2012), 'Data mining concepts and techniques, third edition', Morgan Kaufmann Publishers, Waltham, Mass.
- Kimball R. & Ross M. (2013). The data warehouse toolkit: The definitive guide to dimensional modeling (3rd ed.). John Wiley & Sons.
- Magnitude Types (2023, Φεβρουάριος 5) Ανακτήθηκε από <https://www.usgs.gov/programs/earthquake-hazards/magnitude-types>
- Marketos G., Theodoridis Y. & Kalogeras I. (2008). Seismological Data Warehousing and Mining: A Survey. International Journal of Data Warehousing & Mining, 4(1), 1-16.
- NUTS 2021 level 2 (2023, Μάιος 5) Ανακτήθηκε από <https://ec.europa.eu/eurostat/web/gisco/geodata/reference-data/administrative-units-statistical-units/nuts>
- Özcan, Mert & Peker, Serhat. (2021). Designing a Data Warehouse for Earthquake Risk Assessment of Buildings: A Case Study for Healthcare Facilities. Sakarya University Journal of Computer and Information Sciences. 4. 156-165. 10.35377/saucis.04.01.872729.
- Ponniah P., (2010). Data Warehousing Fundamentals for IT Professionals, Second Edition. John Wiley & Sons.
- PostGIS Geometries (2023, Μάιος 5) Ανακτήθηκε από <http://postgis.net/workshops/postgis-intro/geometries.html>
- Rainardi V., (2008). Building a data warehouse: With examples in SQL Server. Wiley.
- Search for Earthquakes (2023, Ιούνιος 3) Ανακτήθηκε από <https://www.emsc-csem.org/Earthquake/?filter=yes>
- World Health Organization Earthquakes. (2023, Φεβρουάριος 3). Ανακτήθηκε από <https://www.who.int/health-topics/earthquakes>



Inmon W. H., (1996) Building the Data Warehouse. John Wiley & Sons, 1996.

What is a fault and what are the different types? (2023, Φεβρουάριος 4) Ανακτήθηκε από <https://www.usgs.gov/faqs/what-a-fault-and-what-are-different-types>

World Cities (2023, Φεβρουάριος 10) Ανακτήθηκε από <https://hub.arcgis.com/datasets/schools-BE::world-cities/explore>

## Παράρτημα Α: «DDL Αποθήκης δεδομένων»

```
-- Creation of the database
CREATE DATABASE dw_seismological_data
WITH
  OWNER = postgres
  ENCODING = 'UTF8'
  CONNECTION LIMIT = -1
  IS_TEMPLATE = False;

-- Creation of the postgis extension
CREATE EXTENSION postgis;

-- Creation of the schema europe_earthquakes
CREATE SCHEMA europe_earthquakes
  AUTHORIZATION postgres;

-- Setting search path to europe_earthquakes, public
SET SEARCH_PATH TO europe_earthquakes, public;

-- Begin transaction
BEGIN;
-- Creation of the dates dimension
CREATE TABLE dim_dates
(
  date_id varchar(8) PRIMARY KEY NOT NULL,
  year    smallint          NOT NULL,
  month   smallint          NOT NULL,
  day     smallint          NOT NULL,
  season  varchar           NOT NULL
);

-- Create index on dim_dates dimension
CREATE UNIQUE INDEX IF NOT EXISTS idx_dim_dates_date_id ON dim_dates USING
btree (date_id);

-- Creation of the times dimension
CREATE TABLE IF NOT EXISTS dim_times
(
  time_id character varying(6) PRIMARY KEY NOT NULL,
  hour    smallint              NOT NULL,
  minute  smallint              NOT NULL,
  second  smallint              NOT NULL
);

-- Create index for dim_times dimension
CREATE UNIQUE INDEX IF NOT EXISTS idx_dim_times_time_id ON dim_times USING
btree (time_id);

-- Creation of the magnitude types dimension
CREATE TABLE IF NOT EXISTS dim_magnitude_types
(
  magnitude_type character varying(15) PRIMARY KEY NOT NULL,
  description    character varying
);

-- Create index on dim_magnitude_types dimension
```

```
CREATE UNIQUE INDEX IF NOT EXISTS idx_dim_magnitude_types_magnitude_type
ON dim_magnitude_types USING btree (magnitude_type);

-- Creation of the seismic faults dimension
CREATE TABLE IF NOT EXISTS dim_faults
(
    fault_id          character varying PRIMARY KEY NOT NULL,
    fault_geometry    geometry(Polygon, 4326),
    strike_min        double precision,
    strike_avg        double precision,
    strike_max        double precision,
    dip_min           double precision,
    dip_avg           double precision,
    dip_max           double precision,
    rake_min          double precision,
    rake_avg          double precision,
    rake_max          double precision,
    depth_min         double precision,
    depth_max         double precision,
    fault_type        character varying,
    fault_slip        character varying
);

-- Create index on dim_faults dimension
CREATE UNIQUE INDEX IF NOT EXISTS idx_dim_faults_fault_id ON dim_faults
USING btree (fault_id);
-- Create spatial index on dim_faults dimension
CREATE INDEX IF NOT EXISTS sidx_dim_faults_fault_geometry ON dim_faults
USING gist (fault_geometry);

-- Creation of the subductions dimension
CREATE TABLE IF NOT EXISTS dim_subductions
(
    subduction_id     character varying PRIMARY KEY NOT NULL,
    subduction_geometry geometry(Polygon, 4326),
    slab_name         character varying,
    depth_high        double precision,
    depth_low         double precision
);

-- Create index on dim_subductions dimension
CREATE UNIQUE INDEX IF NOT EXISTS idx_dim_subductions_subduction_id ON
dim_subductions USING btree (subduction_id);
-- Create spatial index on dim_subductions dimension
CREATE INDEX IF NOT EXISTS sidx_dim_subduction_subduction_geometry ON
dim_subductions USING gist (subduction_geometry);

-- Creation of the countries dimension
CREATE TABLE IF NOT EXISTS dim_countries
(
    country_id        character varying PRIMARY KEY NOT NULL,
    country_geometry  geometry(MultiPolygon, 4326),
    country_name      character varying
);

-- Create index on dim_countries dimension
CREATE UNIQUE INDEX IF NOT EXISTS idx_dim_countries_country_id ON
dim_countries USING btree (country_id);
-- Create spatial index on dim_countries dimension
```

```
CREATE INDEX IF NOT EXISTS sidx_dim_countries_country_geometry ON
dim_countries USING gist (country_geometry);

-- Creation of the regions dimension
CREATE TABLE IF NOT EXISTS dim_regions
(
    region_id          character varying PRIMARY KEY NOT NULL,
    region_geometry    geometry(MultiPolygon, 4326),
    country_id         character varying REFERENCES dim_countries
(country_id),
    region_name        character varying
);

-- Create index on dim_regions dimension
CREATE UNIQUE INDEX IF NOT EXISTS idx_dim_regions_region_id ON dim_regions
USING btree (region_id);
-- Create spatial index on dim_regions dimension
CREATE INDEX IF NOT EXISTS sidx_dim_regions_region_geometry ON dim_regions
USING gist (region_geometry);

-- Creation of the cities dimension
CREATE TABLE IF NOT EXISTS dim_cities
(
    city_id           smallserial PRIMARY KEY NOT NULL,
    city_geometry     geometry(Point, 4326),
    city_name         character varying,
    region_id         character varying(10) REFERENCES dim_regions (region_id)
);

-- Create index on dim_cities dimension
CREATE UNIQUE INDEX IF NOT EXISTS idx_dim_cities_city_id ON dim_cities
USING btree (city_id);
-- Create spatial index on dim_cities dimension
CREATE INDEX IF NOT EXISTS sidx_dim_cities_city_geometry ON dim_cities
USING gist (city_geometry);

-- Creation of the fact earthquakes
CREATE TABLE IF NOT EXISTS fact_earthquakes
(
    earthquake_id     bigserial PRIMARY KEY NOT NULL,
    date_id           character varying(8) NOT NULL REFERENCES dim_dates
(date_id),
    time_id           character varying(6) NOT NULL REFERENCES dim_times
(time_id),
    subduction_id     character varying REFERENCES dim_subductions
(subduction_id),
    nearest_city_id   integer REFERENCES dim_cities (city_id),
    city_distance      double precision,
    fault_array       character varying[],
    magnitude_type    character varying(15) REFERENCES dim_magnitude_types
(magnitude_type),
    epicenter         geometry(Point, 4326),
    magnitude         double precision,
    hypocenter_depth double precision
);

-- Create index for fact_earthquakes table
CREATE INDEX IF NOT EXISTS idx_fact_earthquakes_id ON fact_earthquakes
USING btree (earthquake_id);
```

```
-- Create spatial index for fact_earthquakes table
CREATE INDEX IF NOT EXISTS sidx_earthquakes_epicenter ON fact_earthquakes
USING gist (epicenter);

-- Commit transaction
COMMIT;

-- This function finds the region of every city
CREATE OR REPLACE FUNCTION fn_cities_region_id()
  RETURNS trigger
  LANGUAGE plpgsql
AS
$$
BEGIN
  new.region_id := (SELECT region_id
                    FROM dim_regions
                    WHERE st_within(NEW.city_geometry, region_geometry)
                    LIMIT 1);

  RETURN new;
END;
$$;

CREATE OR REPLACE TRIGGER trg_cities_region_id
  BEFORE INSERT OR UPDATE OF city_geometry
  ON dim_cities
  FOR EACH ROW
EXECUTE PROCEDURE fn_cities_region_id();

-- This function finds the subduction of every earthquake
CREATE OR REPLACE FUNCTION fn_earthquakes_subduction_id()
  RETURNS trigger
  LANGUAGE plpgsql
AS
$$
BEGIN
  new.subduction_id := (SELECT subduction_id
                       FROM dim_subductions
                       WHERE st_within(NEW.epicenter, subduction_geometry)
                       LIMIT 1);

  RETURN new;
END;
$$;

CREATE OR REPLACE TRIGGER trg_earthquakes_subduction_id
  BEFORE INSERT OR UPDATE OF epicenter
  ON fact_earthquakes
  FOR EACH ROW
EXECUTE PROCEDURE fn_earthquakes_subduction_id();

-- This function finds the nearest city in an earthquake and the distance
CREATE OR REPLACE FUNCTION fn_nearest_city()
  RETURNS trigger
  LANGUAGE plpgsql
AS
$$
BEGIN
  WITH filtered_cities AS (SELECT city_id, city_geometry
                           FROM dim_cities
```

```
WHERE city_geometry &&
st_expand(new.epicenter, 26.1) -- value is in degrees
)
SELECT city_id, st_distancesphere(new.epicenter, city_geometry) / 1000
AS distance
INTO new.nearest_city_id, new.city_distance
FROM filtered_cities
ORDER BY distance
LIMIT 1;
RETURN new;
END;
$$;

CREATE OR REPLACE TRIGGER trg_nearest_city
BEFORE INSERT OR UPDATE OF epicenter
ON fact_earthquakes
FOR EACH ROW
EXECUTE PROCEDURE fn_nearest_city();

-- This function finds all the faults associated with the earthquake
CREATE OR REPLACE FUNCTION fn_earthquake_fault_ids()
RETURNS TRIGGER
LANGUAGE plpgsql
AS
$$
BEGIN
SELECT array_agg(fault_id)
INTO new.fault_array
FROM dim_faults
WHERE st_contains(fault_geometry, new.epicenter);

RETURN new;
END;
$$;

CREATE OR REPLACE TRIGGER trg_earthquake_fault_ids
BEFORE INSERT OR UPDATE OF epicenter
ON fact_earthquakes
FOR EACH ROW
EXECUTE FUNCTION fn_earthquake_fault_ids();
```

## Παράρτημα Β: «Ερωτήματα SQL»

```
SET search_path TO europe_earthquakes, public;
```

```
-----  
-- Query #1 Earthquakes on faults  
-----
```

```
SELECT  
  json_build_object(  
    'type', 'FeatureCollection',  
    'features', json_agg(  
      json_build_object(  
        'type', 'Feature',  
        'id', earthquake_id,  
        'geometry', ST_AsGeoJSON(epicenter)::json,  
        'properties', json_build_object()  
      )  
    )  
  ) AS geojson  
FROM fact_earthquakes  
WHERE fault_array IS NOT NULL;
```

```
-----  
-- Query #2 Earthquakes on subductions  
-----
```

```
SELECT  
  json_build_object(  
    'type', 'FeatureCollection',  
    'features', json_agg(  
      json_build_object(  
        'type', 'Feature',  
        'id', earthquake_id,  
        'geometry', ST_AsGeoJSON(epicenter)::json,  
        'properties', json_build_object()  
      )  
    )  
  ) AS geojson  
FROM fact_earthquakes  
WHERE subduction_id IS NOT NULL;
```

```
-----  
-- Query #3 10 Greatest Earthquakes near Patra the last decade  
-----
```

```
SELECT  
  json_build_object(  
    'type', 'FeatureCollection',  
    'features', json_agg(  
      json_build_object(  
        'type', 'Feature',  
        'id', earthquake_id,  
        'geometry', ST_AsGeoJSON(epicenter)::json,  
        'properties', json_build_object(  
          'earthquake_id', earthquake_id,  
          'magnitude', magnitude,  
          'date', date,  
        )  
      )  
    )  
  ) AS geojson  
FROM fact_earthquakes  
WHERE date >= '2010-01-01' AND date <= '2019-12-31';
```

```

        'season', season,
        'time', time,
        'city', city
    )
)
) AS geojson
FROM (
SELECT
    fe.earthquake_id,
    fe.epicenter,
    fe.magnitude,
    dd.year || '/' || dd.month || '/' || dd.day AS date,
    dd.season,
    dt.hour || ':' || dt.minute || ':' || dt.second AS time,
    dc.city_name AS city
FROM fact_earthquakes fe
    INNER JOIN dim_cities dc ON dc.city_id = fe.nearest_city_id
    INNER JOIN dim_dates dd ON dd.date_id = fe.date_id
    INNER JOIN dim_times dt ON dt.time_id = fe.time_id
WHERE nearest_city_id = 51
    AND dd.year >= 2014
ORDER BY magnitude DESC
LIMIT 10
) AS subquery;

-----
-- Query #4 Percentage of earthquakes per country
-----

WITH cities_regions_countries AS (
SELECT
    dc.city_id,
    dc.city_geometry,
    dc.city_name,
    dr.region_name,
    dr.region_geometry,
    dcn.country_id,
    dcn.country_geometry,
    dcn.country_name
FROM
    dim_regions dr
    INNER JOIN dim_cities dc ON dr.region_id = dc.region_id
    INNER JOIN dim_countries dcn ON dcn.country_id = dr.country_id
)

SELECT
    json_build_object(
        'type', 'FeatureCollection',
        'features', json_agg(
            json_build_object(
                'type', 'Feature',
                'id', country_id,
                'geometry', country_geometry,
                'properties', json_build_object(
                    'country_id', country_id,
                    'country', country,
                    'total_earthquakes', total_earthquakes,
                    'percentage', "percentage %"
                )
            )
        )
    )

```



```

        )
    )
) AS geojson
FROM (
    SELECT
        crc.country_id AS country_id,
        crc.country_geometry AS country_geometry,
        crc.country_name AS country,
        COUNT(*) AS total_earthquakes,
        ROUND((COUNT(*) / (SUM(COUNT(*) OVER ())) * 100, 0) AS "percentage
%"
    FROM
        fact_earthquakes fe
        INNER JOIN cities_regions_countries crc ON crc.city_id =
fe.nearest_city_id
    GROUP BY
        crc.country_id, crc.country_geometry, crc.country_name
    ORDER BY
        COUNT(*) DESC
) AS subquery;

```

```

-----
-- Query #5 Percentage of faults per country
-----

```

```

WITH cities_regions_countries AS (
    SELECT
        dc.city_id,
        dc.city_geometry,
        dc.city_name,
        dr.region_name,
        dr.region_geometry,
        dcn.country_id,
        dcn.country_geometry,
        dcn.country_name
    FROM
        dim_regions dr
        INNER JOIN dim_cities dc ON dr.region_id = dc.region_id
        INNER JOIN dim_countries dcn ON dcn.country_id = dr.country_id
)
SELECT
    json_build_object(
        'type', 'FeatureCollection',
        'features', json_agg(
            json_build_object(
                'type', 'Feature',
                'id', country_id,
                'geometry', country_geometry,
                'properties', json_build_object(
                    'country_id', country_id,
                    'country', country,
                    'total_faults', total_faults,
                    'percentage', "percentage %"
                )
            )
        )
    )
)

```

```

) AS geojson
FROM (
  SELECT
    crc.country_id AS country_id,
    crc.country_geometry AS country_geometry,
    crc.country_name AS country,
    COUNT(df.fault_id) AS total_faults,
    ROUND((COUNT(df.fault_id) / (SUM(COUNT(df.fault_id)) OVER ())) *
100, 0) AS "percentage %"
  FROM
    fact_earthquakes fe
    INNER JOIN cities_regions_countries crc ON crc.city_id =
fe.nearest_city_id
    INNER JOIN dim_faults df ON df.fault_id = ANY(fe.fault_array)
  GROUP BY
    crc.country_id, crc.country_geometry, crc.country_name
  ORDER BY
    COUNT(df.fault_id) DESC
) AS subquery;

-----
-- Query #6 Top 20 cities with most earthquakes
-----

WITH cities_regions_countries AS (
  SELECT
    dc.city_id,
    dc.city_geometry,
    dc.city_name,
    dr.region_name,
    dr.region_geometry,
    dcn.country_id,
    dcn.country_geometry,
    dcn.country_name
  FROM
    dim_regions dr
    INNER JOIN dim_cities dc ON dr.region_id = dc.region_id
    INNER JOIN dim_countries dcn ON dcn.country_id = dr.country_id
)

SELECT
  json_build_object(
    'type', 'FeatureCollection',
    'features', json_agg(
      json_build_object(
        'type', 'Feature',
        'id', city_id,
        'geometry', city_point,
        'properties', json_build_object(
          'city_id', city_id,
          'city', city,
          'country', country,
          'total_earthquakes', total_earthquakes
        )
      )
    )
  ) AS geojson
FROM (
  SELECT

```

```

        fe.nearest_city_id AS city_id,
        crc.city_name AS city,
        crc.city_geometry AS city_point,
        crc.country_name AS country,
        COUNT(fe.nearest_city_id) AS total_earthquakes
    FROM
        fact_earthquakes fe
        INNER JOIN cities_regions_countries crc ON crc.city_id =
fe.nearest_city_id
    GROUP BY
        fe.nearest_city_id,          crc.city_name,          crc.city_geometry,
        crc.country_name
    ORDER BY
        COUNT(fe.nearest_city_id) DESC
    LIMIT 20
) AS subquery;

```

```

-----
-- Query #7 Top 20 earthquakes
-----

```

```

WITH cities_regions_countries AS (
    SELECT
        dc.city_id,
        dc.city_geometry,
        dc.city_name,
        dr.region_name,
        dr.region_geometry,
        dcn.country_id,
        dcn.country_geometry,
        dcn.country_name
    FROM
        dim_regions dr
        INNER JOIN dim_cities dc ON dr.region_id = dc.region_id
        INNER JOIN dim_countries dcn ON dcn.country_id = dr.country_id
)

```

```

SELECT
    json_build_object(
        'type', 'FeatureCollection',
        'features', json_agg(
            json_build_object(
                'type', 'Feature',
                'id', earthquake_id,
                'geometry', ST_AsGeoJSON(epicenter)::json,
                'properties', json_build_object(
                    'date', date,
                    'season', season,
                    'time', time,
                    'city', city,
                    'country', country,
                    'earthquake_id', earthquake_id,
                    'magnitude', magnitude
                )
            )
        )
    ) AS geojson
FROM (
    SELECT

```

```

dd.year || '/' || dd.month || '/' || dd.day AS date,
dd.season AS season,
dt.hour || ':' || dt.minute || ':' || dt.second AS time,
crc.city_name AS city,
crc.country_name AS country,
fe.earthquake_id AS earthquake_id,
fe.epicenter AS epicenter,
fe.magnitude AS magnitude
FROM
    fact_earthquakes fe
    INNER JOIN cities_regions_countries crc ON crc.city_id =
fe.nearest_city_id
    INNER JOIN dim_dates dd ON dd.date_id = fe.date_id
    INNER JOIN dim_times dt ON dt.time_id = fe.time_id
ORDER BY
    fe.magnitude DESC
LIMIT 20
) AS subquery;

```

-----  
-- Query #8 Earthquakes percentage on faults/subductions  
-----

```

SELECT COUNT(subduction_id) AS in_subduction
    , ROUND((COUNT(subduction_id) / (SUM(COUNT(*)) OVER ())) * 100, 0)
AS "subduction %"
    , COUNT(fault_array) AS in_fault
    , ROUND((COUNT(fault_array) / (SUM(COUNT(*)) OVER ())) * 100, 0) AS
"fault %"
FROM fact_earthquakes fe;

```

-----  
-- Query #9 Total and Top earthquakes per year  
-----

```

SELECT dd.year AS year
    , COUNT(fe.magnitude) AS total
    , ROUND((COUNT(fe.magnitude) / (SUM(COUNT(*)) OVER ())) * 100, 0) AS
"percentage %"
    , MAX(fe.magnitude) AS max
FROM fact_earthquakes fe
    INNER JOIN dim_dates dd ON dd.date_id = fe.date_id
GROUP BY 1;

```

-----  
-- Query #10 Total and Top earthquakes per month  
-----

```

SELECT dd.month AS month
    , COUNT(fe.magnitude) AS total
    , ROUND((COUNT(fe.magnitude) / (SUM(COUNT(*)) OVER ())) * 100, 0) AS
"percentage %"
    , MAX(fe.magnitude) AS
max

```

```
FROM fact_earthquakes fe
      INNER JOIN dim_dates dd ON dd.date_id = fe.date_id
WHERE dd.year > 2004
      AND dd.year < 2023
GROUP BY 1;
```

```
-----
-- Query #11 Total and Top earthquakes per season
-----
```

```
SELECT dd.season AS season
      , COUNT(fe.magnitude) AS total
      , ROUND((COUNT(fe.magnitude) / (SUM(COUNT(*)) OVER ())) * 100, 0) AS
"percentage %"
      , MAX(fe.magnitude) AS max
FROM fact_earthquakes fe
      INNER JOIN dim_dates dd ON dd.date_id = fe.date_id
WHERE dd.year > 2004
      AND dd.year < 2023
GROUP BY 1;
```

Υπεύθυνη Δήλωση Συγγραφέα:

Δηλώνω ρητά ότι, σύμφωνα με το άρθρο 8 του Ν.1599/1986, η παρούσα εργασία αποτελεί αποκλειστικά προϊόν προσωπικής μου εργασίας, δεν προσβάλλει κάθε μορφής δικαιώματα διανοητικής ιδιοκτησίας, προσωπικότητας και προσωπικών δεδομένων τρίτων, δεν περιέχει έργα/εισφορές τρίτων για τα οποία απαιτείται άδεια των δημιουργών/δικαιούχων και δεν είναι προϊόν μερικής ή ολικής αντιγραφής, οι πηγές δε που χρησιμοποιήθηκαν περιορίζονται στις βιβλιογραφικές αναφορές και μόνον και πληρούν τους κανόνες της επιστημονικής παράθεσης.