Homework 1 & 2 for DMDS exam project

Professors: Prof. Rosati, Prof. Lembo.

Project members: Felli Stefano (1896877), Cirillo Lorenzo (1895955).

Database Management system: PostgreSQL.

Chosen Database: Countries-States-Cities (https://www.kaggle.com/datasets/darshangada/countries-states-cities-database?select=csv)

The database contains a collection of data providing detailed geographical information about cities, states, countries, continents, and sub continents, each one modeled as a table.

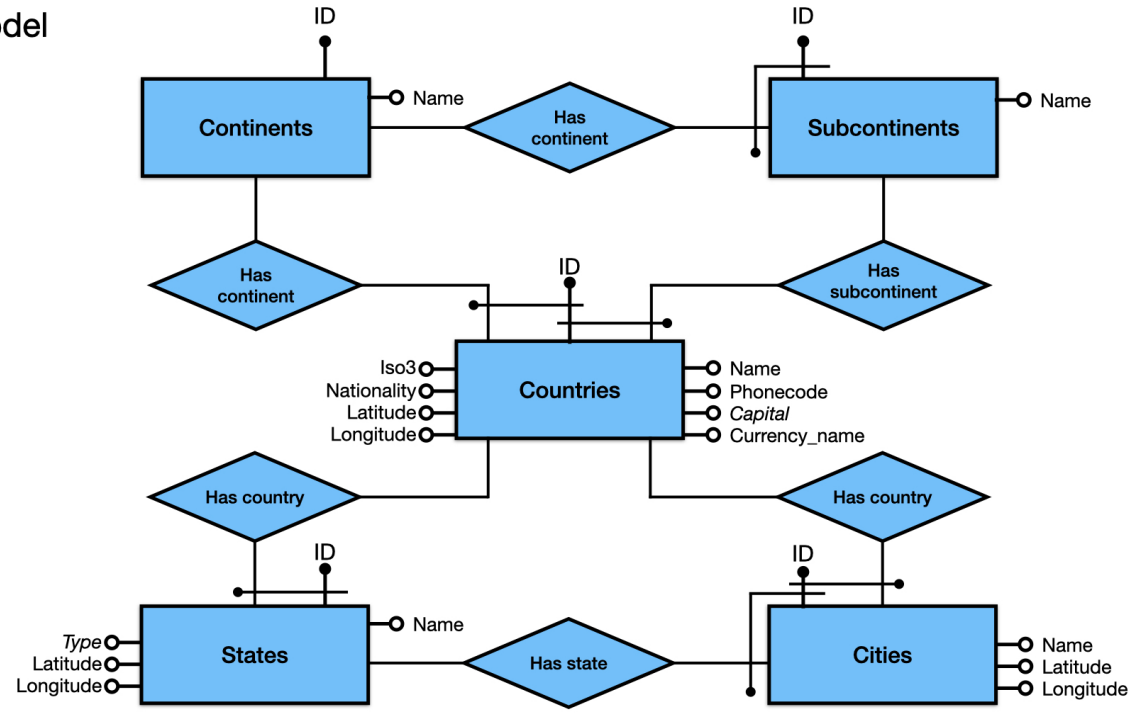
There are 155.935 entries in total, broken down as follows:

- Total Continents : 6
- Total Sub Continents : 22
- Total Countries : 250
- Total States/Regions/Municipalities/Provinces : 5,084
- Total Cities/Towns/Districts : 150,573

Entity Relationship model

The following E-R model reports the structure and all the constraints of the database

E-R Model



For each table, the attributes and the constraints are resumed:

- CONTINENTS(**id**, name)

- SUBCONTINENTS(**id**, name, continent_id)
- COUNTRIES(**id**, name, iso3, phonecode, *capital*, currency_name, *continent_id*, *subcontinent_id*, nationality, latitude, longitude)
- STATES(**id**, name, country_id, *type*, latitude, longitude)
- CITIES(**id**, name, state_id, country_id, latitude, longitude)

N.B. black dotted/bold attributes are primary keys, italic attributes can assume NULL values.

Get Started

- run `/2_preprocess.py` (to modify the original csv files)

In PostgreSQL:

- run `/3_database_definition.sql` (to create tables and set constraints on tables)
- run `/4_non_optimized_queries.sql` (to run the non optimized queries)
- run `/5_optimized_queries.sql` (to run the optimized queries and their unefficient version)