Project Final Report

INST327

Course Section: _ESG1_ Team Number/Name: _____#7_____

Date: __5/14/2021____

This is the team 7 project final report and is divided into the following sections:

- Introduction - project description, scope and use cases
- Database Description - logical design, physical database, sample data, views and queries.
- Changes from Original Design - including entity relationships, questions to answer, data sources
- Lessons Learned - what would we do differently?
- Potential Future Work - how can we make it better?

1. Introduction

The project idea that we have decided to go with was Global warming and the data that is associated with global warming. Global warming is a huge problem in the world and has been attributed to rising sea levels and extreme weather events that affected our daily lives. Being able to decipher the data, we would be able to determine the severity that it has on the environment and other factors in the world. Many people believe that global warming will eventually cause irreversible damage to countries and places all over the world, so if there is a way to determine the severity of it and the effects it causes, it would be very beneficial to see.

Some of the data that we'll be directly looking at is global temperature history, greenhouse gas emotions, sea level history, deforestation, premature deaths, and wildfires. These data sets have certain information such as historical information that includes trends, the statistical values

of things like CO2 levels and temperature, and the statistical values of temperature in certain regions all around the world. We will also have some demographic information about countries outside of the US and how global warming has affected the environment and temperatures. With the usage of both US and other countries around the world, it would allow us to create correlations between countries all around the world and global warming data.

Ethics were not a big concern when designing the database since there are no real privacy problems. The main ethical issue would be data integrity so good decision making is possible.

The use cases of this database are to answer some of the following questions.

- Who is contributing the most to greenhouse gases?
- How fast will the sea levels rise based on history and the amount of greenhouse gases?
- How fast will deforestation be based on history and the amount of greenhouse gases?
- How does the relationship between deforestation and wildfire affect global warming?
- How has global temperature history and greenhouse gas emissions history affect the life span in different nations across time?
- What is the relationship between the increase in global temperature history on the amount of wildfires that occur in the US?

2. Database Description

Below the logical design, physical database, description of the sample data and sources, and some of the views used to answer the questions above.

2.1 Logical Design

Below are logical areas of the database tables and the Entity Relationship Diagram (ERD). There are three basic logical areas of the entity relationship diagram, country demographics, causes, and effects. The country demographics contain the demographic data like population, area, and life expectancy, the causes contain the causes of global warming and the effects contain the effects of global warming.

· Countries/continents demographics

- countries table – description of countries
- continents table – description of continents
- continents_countries table – historical information for countries

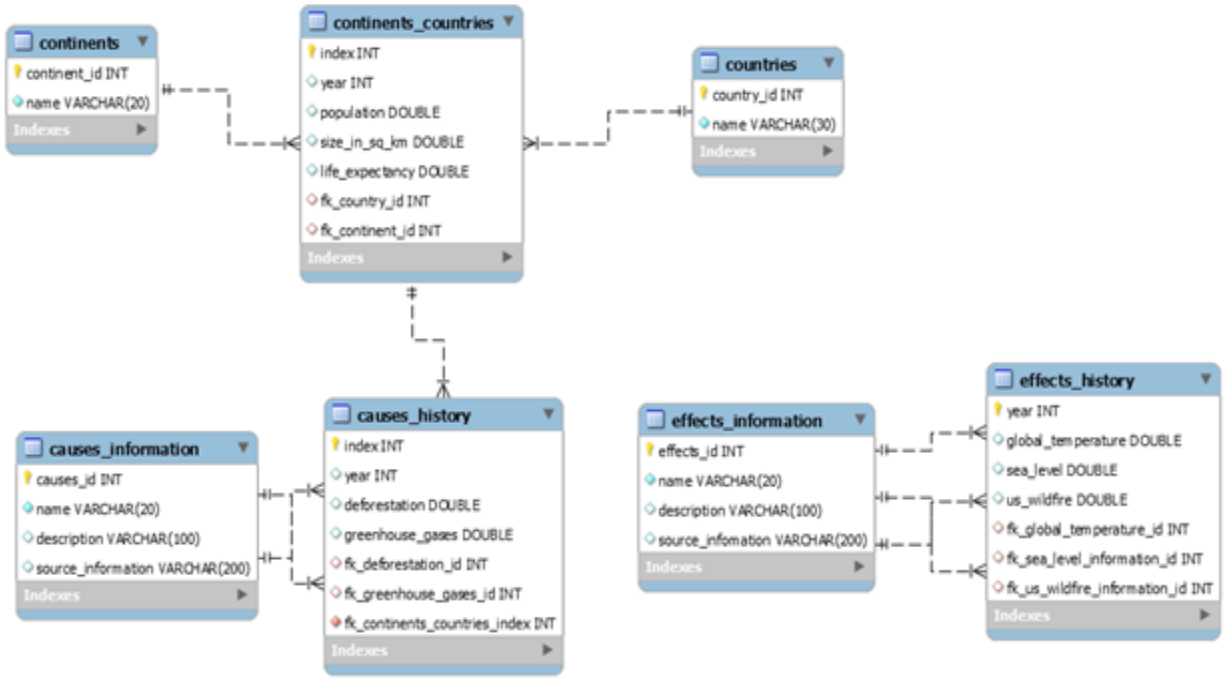
· Causes of global warming

- causes_history table – historical causes data
- causes_information table – causes description information

· Effects of global warming

- effects_history table – historical effects data
- effects_information table – effects description information

There is a one-to-many relationship between the countries and the continents_countries data which is a linking table. The other side of the link is between the continents and the continents_countries data. There is a one-to-one relationship between continents_countries data and the causes_history since we have data for 10 different countries. The causes_information has a one-to-many relationship with causes_history. The effects_information table has a one-to-many relationship with effects_history. The Entity Relationship Diagram is shown below.



2.2 Physical Database

The physical database is made up of seven tables. Below are the physical tables in the global warming database along with the primary and foreign keys and descriptions of the columns in each table. It was created directly from the ERD and was straight forward.

cause_history - the yearly values of the causes like deforestation and greenhouse gases

- Primary key : index to row
- year – year data gathered for country
- deforestation – how many acres of land deforested in country
- greenhouse_gases – how much greenhouse gases produced by country
- fk_deforestation_id – foreign key to deforestation causes information table
- fk_greenhouses_gases_id – foreign key to greenhouse gases information
- fk_continents_countries_index – foreign key to continent/country table index

causes_information - the reference information of the causes of global warming

- Primary key: causes_id
- name – name of the cause
- description – includes detail and units
- source_information – includes URL to access raw data

continents- the continents

- primary key: continent_id
- name - name of continent

continents_countries – contains the demographic information history for countries

- primary key: index to row
- year - year of data gathered
- population - country population count
- size_in_sq_km
- life_expectancy - in years
- fk_country_id - foreign key to the countries table
- fk_continent_id - foreign key to the continents table

countries – names of the countries

- primary key: country_id
- name - name of country

effects_history – history of the effects of global warming

- primary key : year
- global_temperature - in celsius
- sea_level - relative to year 2000
- us_wildfire - acres burned
- fk_global_temperature_id - foreign key to global temperature information
- fk_sea_level_information_id - foreign key to sea level information

- fk_us_wildfire_information_id - foreign key to wildfire information

effects_information- reference information of the effects of global warming

- primary key: effects_id
- name - name of effect
- description - units and detail information about effect
- source_ information - URL to access information

2.3 Sample Data

Sample data was retrieved from the internet at the websites shown in the table below to populate the database tables. For the country specific data we selected the top ten greenhouse gas producers as shown at this site, https://www.ucsusa.org/resources/each-countrys-share-co2-emissions. These countries are China, United States, Germany, Japan, India, Russion Federation, Iran, Saudi Arabia, South Korea, and Indonesia. The data on the websites was provided in CSV format so it was just a matter of downloading the data and copy and pasting into a new an SQL file to create the rows or in most cases create a CSV file to be imported into the database. We have a few tables that do not meet the 15 count because there are only 7 continents, and the causes and effects tables only have three or fours records since there is one record for each information reference as shown in the table below.

Website	Data retrieved	Database table(s)
https://www.ncdc.noaa.gov/cag/global/time-series/globe/land/12/12/1990-2019	Global temperature from 1990 to 2019	effects_history, effects_information
https://www.climate.gov/sites/default/files/Climate_dot_gov_dashboard_SeaLevel_Jan2021update.txt	Sea level history from 1990 to 2019	effects_history, effects_information
https://www.nifc.gov/fire-information/statistics/wildfires	US forest fires from 1990 to 2020	effects_history, effects_information
https://www.globalforestwatch.org/dashboards	Top ten countries forest cover loss from 2001 to 2020	causes_history, causes_information

https://raw.githubusercontent.com/owid/co2-data/master/owid-co2-data.csv	Top ten countries greenhouse gas emission from 1990 to 2019	causes_history, causes_information
https://data.worldbank.org/country	Top ten countries demographic information, population, area, life expectancy from 1990 to 2019	continents_countries table

2.4 Views and Queries

We created five queries to satisfy the final project requirement. We also ran a query on each table to verify the sample data was correct. Below are the queries listed and the requirements table.

- view1_country_avg_population.sql
- view2_2018_global_temp_and_description.sql
- view3_max_greenhouse_gases_by_year.sql
- view4_population_sea_level_history.sql
- view5_country_info.sql

View Name	Req. A	Req. B	Req. C	Req. D	Req. E
view1_country_avg_population.sql			X		
view2_2018_global_temp_and_description.sql	X	X			
view3_max_greenhouse_gases_by_year.sql	X	X	X		
view4_population_sea_level_history.sql	X		X		
view5_country_info.sql	X	X		X	X

3. Changes from original design

Below are a list of the changes from the original design. They include changes to the ERD to make it more normalized and changes to make it easier to manage the sample data.

- Removed the taxes and patent tables because they were not pertinent to what we were trying to build, and it was pointed out in the project review.
- Created a continent/countries table that was a linking table between the countries table and the continents table. It contained the demographic history information.
- We limited the countries to the top ten greenhouse gas producers because the sample data was getting large to manage. Also, the top ten accounts for 68% of the greenhouse gases.
- We added to the continent/countries table demographic information like population, area in square kilometers, and life expectancy which could be causes and effects for global warming.
- The ERD review had us normalize the tables by moving the causes information, effects information to their own table. This was also done for the countries and continents.

4. Lessons Learned

The biggest lesson learned is to start early and save enough time to validate the sample data that is retrieved from the websites.

5. Potential Future Work

There are several areas for future work.

- Validate the sample data and repopulate the tables with correct data.
- Add more countries' information to get a better picture of causes and effects.
- Create new queries to answer the questions about causes and effects.
- Add weather events to effects history, like hurricanes and typhoons.
- Get individual country data on the effects like sea level and weather events.