

Assignment 4: Dynamically Generating Insert Statements

1. Dataset file CSVs (also attached in assignment)
 - a. Flow data - Middle_Vernon_Creek_Hydro_Data.csv
 - b. Wildfire data - wildfire_data_bc.csv
 - c. Weather data - Kelowna_48369_station_data.csv
2. Updated ER diagram
 - a. Diagram: Please see at bottom of this document
 - b. Changes made and justification:
 - Deleted Season entity because it makes more logical sense to use annual averages for precipitation and water flow data considering wildfire data is given as annual hectares of fire. We are interested in year to year correlations between the data, not seasonal differences.
 - Changed attributes of Precipitation to only total participation in mm. This simplifies rainfall and snowfall which are in different units and aligns better with our exact data. We are using one specific location (Kelowna Airport) as our weather station of interest.
 - Location attribute of river flow changed to foreign key
 - Occurs in relationship removed between wildfire and location as the wildfire data is general to all of BC
 - Changed Weather Data entity attributes to reflect our weather data csv file. We decided it was easier to classify dates by year, month, day rather than time stamp because we don't care about hourly changes.
 - Updated diagram to reflect that Weather Station ID is the foreign key for Weather Data and Precipitation ISA Weather Data so ID also functions as its foreign key.
 - Grouped the River Flow date attributes on unique (month, year) combinations since the data file was otherwise too large. These month and year combinations can be used together with the precipitation data.
 - Added additional attributes to Wildfire entity that will be useful in data analysis including number of fires and max area (largest fire).
 - Included Location ID as an attribute of location as a primary key in case there are multiple locations with very similar names.

The ISA relationship is a total relationship. Every Precipitation entity is a Weather Data entity. We chose to separate the two because we felt that it better represents the fact that it is specifically precipitation data that has a relation to wildfire data rather than Weather Data.

The Location entity was revised to provide more attributes. The Weather Station was kept separate with its own ID in order to represent the slight difference in locations between our hydrological, weather, and wildfire data.

3. Relational schema

WeatherStation(ID)
Location(Name, Latitude, Longitude)
WeatherData(LocalYear, LocalMonth, LocalDay, MeanTemperature, ID)
Precipitation(TotalPrecipitation, ID)
Riverflow(LocalYear, LocalMonth, AverageFlow, Location)
Wildfire(Year, NumberOfFires, TotalArea, MaxArea)

4. Python file

Please see attached.

5. SQL DDL file

Please see attached.

