**SQL – Final Project – Valerie Briot**

For this assignment, we will consider databased called BuildingEnergy that will be used to model the energy usage of buildings in NYC.

**Data Modeling:**

1. **Entities:**
   1. EnergyCategories: entity to store categories of energy types.
   2. EnergyTypes: entity to store types of energy, each type of energy will belong to only one category.
   3. Buildings: entity to store buildings. Currently we are focusing only on building in NYC but the model is set-up to accommodate multiple cities/countries. Building belongs to only one city.
   4. Cities: entity to store cities to which building are in. A city will have multiple building. We will assume that a city has belongs to only one country. Currently the only city under consideration will be NYC.
   5. Countries: entity to store countries. Country will have multiple cities. The only country we are currently considering is “United States of America”. The attribute: country\_sequencing is an arbitrary sequence number to allow for a custom sort sequence. For example, United State will be 1.
2. **Relationships**
   1. EnergyCategories to EnergyTypes: One to Many. Note: an energy category must have at least one energy type otherwise it will not be considered and entered in Database. An Energy Type must be classified in a given category.
   2. Cities to Buildings: One to Many. Note: We will not consider a city unless we have a building in it. Also, a building must be in a given city.
   3. Countries to Cities: One to Many. Note: a country will not be entered in DB unless we have cities/buildings that we are considering. A city must belong to a given country.
   4. Buildings to EnergyTypes: Many to Many. A building can use multiple energy type at the same time and in turn, we would expect that same energy type will be used in different building. To model this relationship, an associate table Buildings\_EnergyTypes\_Usage.

**Table Definitions:**

The following are the table definitions that will be entered in the database:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Table = EnergyCategories |  |  |  |  |
| Physical Name | **Data Type** | **Required?** | **PK** | **Notes** |
| energy\_category\_id | INTERGER | Yes | Yes | Uniquely Identify an energy category |
| energy\_category\_code | CHAR(1) | Yes |  | 1 character code that indicates the energy category |
| energy\_category\_description | VARCHAR(25) | Yes |  | Descriptive name of the Energy Category |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Table = EnergyTypes |  |  |  |  |
| Physical Name | **Data Type** | **Required?** | **PK** | **Notes** |
| energy\_type\_id | INTEGER | Yes | Yes | Uniquely Identify an energy type |
| energy\_type\_code | CHAR(2) | Yes |  | 2 character code that indicates Energy Type |
| energy\_type\_description | VARCHAR(25) | Yes |  | Descriptive name of the Energy Type |
| energy\_category\_id | INTEGER | Yes |  | Energy Category of the Energy Type |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Table = Countries |  |  |  |  |
| Physical Name | **Data Type** | **Required?** | **PK** | **Notes** |
| country\_id | INTEGER | Yes | Yes | Uniquely Identify a Country |
| country\_code | CHAR(2) | Yes |  | 2 character code that indicates country |
| country\_name | VARCHAR(50) | Yes |  | Name of country |
| country\_sequencing | INTEGER | Yes |  | Sequence number to allow different sort sequence |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Table = Cities |  |  |  |  |
| Physical Name | **Data Type** | **Required?** | **PK** | **Notes** |
| city\_id | INTEGER | Yes | Yes | Uniquely Identify a city |
| city\_name | VARCHAR(50) | Yes |  | Name of city |
| country\_id | INTEGER | Yes |  | Country id of country city resides in |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Table = Buildings |  |  |  |  |
| Physical Name | **Data Type** | **Required?** | **PK** | **Notes** |
| building\_id | INTERGER | Yes | Yes | Uniquely Identify a building |
| building\_name | VARCHAR(50) | Yes |  | Name of Building |
| city\_id | INTERGER | Yes |  | City id of city where the building resides in |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Table = Building\_EnergyTypes\_Usage | |  |  |  |
| Physical Name | **Data Type** | **Required?** | **PK** | **Notes** |
| building\_energy\_type\_usage\_id | INTEGER | Yes | Yes | Uniquely Identify an energy type currently being used in a given building |
| building\_id | INTEGER | Yes |  | Unique id of the Building |
| energy\_type\_id | INTEGER | Yes |  | Unique id of the Energy Type being used in the Building |

**ER Diagram:**

(ER Diagram constructed using Microsoft Office Visio Professional 2003)



**Data Sets:**

**Table = EnergyCategories:**

|  |  |  |
| --- | --- | --- |
| energy\_category\_id | engergy\_category\_code | energy\_category\_description |
| 1 | F | Fossil |
| 2 | R | Renewable |

**Table = EnergyTypes:**

|  |  |  |  |
| --- | --- | --- | --- |
| energy\_type\_id | energy\_type\_code | energy\_type\_description | energy\_category\_id |
| 1 | EL | Electricity | 1 |
| 2 | FO | Fuel Oil | 1 |
| 3 | GS | Gas | 1 |
| 4 | SL | Solar | 2 |
| 5 | ST | Steam | 1 |
| 6 | WD | Wind | 2 |
| 7 | GT | Geothermal | 2 |

**Table = Countries:**

|  |  |  |  |
| --- | --- | --- | --- |
| country\_id | country\_code | country\_name | country\_sequencing |
| 1 | US | United States of America | 1 |

**Table = Cities:**

|  |  |  |
| --- | --- | --- |
| city\_id | city\_name | country\_id |
| 1 | New York | 1 |

**Table = Buildings:**

|  |  |  |
| --- | --- | --- |
| building\_id | building\_name | city\_id |
| 1 | Borough of Manhattan Community College | 1 |
| 2 | Chrysler Building | 1 |
| 3 | Empire State Building | 1 |
| 4 | Bronx Lions House | 1 |
| 5 | Brooklyn Childrens Museum | 1 |

**Table = Buildings\_EnergyTypes\_Usage:**

|  |  |  |
| --- | --- | --- |
| building\_energy\_type\_usage\_id | building\_id | energy\_type\_id |
| 1 | 1 | 4 |
| 2 | 1 | 1 |
| 3 | 1 | 5 |
| 4 | 2 | 5 |
| 5 | 2 | 1 |
| 6 | 3 | 1 |
| 7 | 3 | 5 |
| 8 | 3 | 3 |
| 9 | 4 | 7 |
| 10 | 5 | 1 |
| 11 | 5 | 7 |

**Queries and SQL Statements:**

See SQL script:

**Future DB development - Tracking Changes over time:**

In this assignment, we have not taken into consideration whether any of the relationship would change over time or whether any of the entry in each entity would change over time. Let us do so now.

**Countries/Cities:**

Even though it is possible that cities and even countries be impacted over time, cities being renamed, changing country they resides in, or even being destroyed, countries being renamed or dissolved. The kind of turmoil such changes are indicators of are being the scope of this analysis. Therefore we will assume that countries/Cities do not need to be tracked overtime.

**EnergyCateories/EnergyTypes:**

Over time it may be possible that an energy type change category, Scientific progress may lead to an energy type being now produced on renewable basis (for example, Electricity could be produced using renewable mean). Hence the energy type should be categorized taking into consideration that the same energy type could have different categories. Taking into considerations these factors, I would propose the following changes to the data model:

1. Create a Sub-Energy type that will be associated with a category (and only one)
2. Energy type could have 1 or more subtypes.
3. Building energy need would be tracked at the energy subtype level ton ensure that the energy category

**Buildings:**

A building life cycle will most probably spans many years but eventually a building may be demolished to make room for new construction. Hence, we should track the start date and end date of a building.

Start date: date building construction is done

End date: date building is demolished

Other attributes may be of interest in an energy study; size of building, materials used for constructions, occupancy level, commercial or residential, …

**Building\_EnergyTypes\_Usage:**

As per discussion above, this relation would now be with EnergySubTypes. In this relation, we should definitely keep tract of time element. A start date/end date pair should be on record to indicate the time duration that this energy subtype was used for a particular building. We would have to validate that the start date/end date are within the date interval of the building itself using check constraint at the column level.

See attached Modified ER Diagram:



**Possible Report with the Time Element:**

Ideally, we would have another relationship “consumption” that would track how much energy is used by the building. This entity would be a child entity of the remodeled “building\_energy\_type\_usage” and would have a granularity of year/month, it would also have a measure of energy used/spent. Each record would be usage for each energy subtype for a month per building. We would expect as time progress, the usage would shift from non-renewable to renewable energy. Once we have captured the data to this granularity we can aggregate it as required.

)

**Mockup Screens:**

(not done, sorry, ran out of time)