**CS443 - Lab 3**

**Question 1:**

Suppose that our database has the following table.

Person

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Con  ID | Con  Name | Con  Pop | Con  Size | State  code | State  Name | State  Rgn | State  Size | State  Pop | Cty  Code | Cty  Name | Cty  Size | Per  SSN | Per  Name | Per  Age | Per  DofB | Per  Add |

**Field Explanation:**

Con: Stands for Country

Pop: Stands for population

Rgn: Stands for region (like west, east, central, etc.)

Cty: Stands for City

Per: Stands for Person

DofB: Stands for date of birth

Add: Stands for Address

**It is assumed that**

* Every country in the world has a different country ID
* Every city in the world has a different city code
* Every state in the world has a different state code, and
* Every person in the world has a different SSN
* Every person in the world has only one citizenship and has only one address

**Other Assumptions:**

* There is no village, county, area, etc. A country consists of several states and each state has several cities

1. **Based on the above assumptions, what do you choose to be the primary key of Person table? Why?**

For the person table, the primary key will be PerSSN. That is because a person is unique to their own SSN and no other person can match a person’s SSN. With the other following attributes, they would either be insignificant to the table or redundant. The following include attributes from Con, State, or City since it does not reflect off the Person table and since it is a Person table, those attributes cannot be chosen as a primary key. As of for the other attributes of Per, those keys cannot be unique since there is a probability that in those attributes it will match within it (i.e., PerAge – 14, 15, 17, 14, cannot be chosen since 14 shows up twice.)

1. **Explain the anomalies exist in the Person table. Choose only one example of insert anomaly, one example of delete anomaly and one example of update anomaly. Note that update does not mean adding or deleting records. It only refers to modifications of values in some rows of the table.**

**Insert Anomaly:**

* Inserting any attribute that is not related to the Person table.
* That includes Con attributes, State attributes, and City attributes because if they were to be included in the Person table, it is needed to have one person to be apart of the following Con, State, and City attributes.
* For example, to have a city, it must be important that one person resides for it to access other attributes within the city.

**Delete Anomaly:**

* Deleting any attribute on the table affects all other attributes that are from the right of that attribute, or in other words, it will be remove what is with that chosen attribute to not leave any gaps within the table.
* Examples shown below:
* PerSSN
  + Deleting this attribute will cause problems for the whole Person table because it is the primary key of that table.
* CtyCode
  + Deleting this portion will discard everything that’s from CtyCode to right, or in other terms, there will be no record of what is in that table. The consequences of that are if a person resides in that city, the person will be technically homeless or have no place to be since that person’s record has been lost for good.
* ConSize
  + Same result as CtyCode and deleting that will cause others within that country to be in trouble or unrecognized. Deleting the following attributes will cause even more greater consequences and deleting that much will be troublesome for the following attributes that are of the left since some of the attributes are dependent off the deleted attributes.

**Update Anomaly:**

* ConPop
  + Since the population of the country can change and is dependable by the StatePop, it will be constantly changing, and it is not possible to control the population that is nationwide.
  + The ConPop always changes over time and would not stay constant or the same, thus making it inconsistent.
* StatePop
  + Similar to ConPop, the StatePop is what makes up the ConPop as a whole.
  + Without one of the StatePop that is inside of the country, it would be difficult to know what has occurred.
  + It will also be inconsistent as ConPop since ConPop is dependable to the StatePop.

1. **Normalize the table; create as many as tables necessary such that all new tables are in third normal form. All the transitive and derived dependencies must be removed..**

**Passes both first and second normal form, but not third normal form.**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Person** | **PerSSN** | **PerName** | **PerAge** | **PerDofB** | **PerAdd** |

|  |  |  |  |
| --- | --- | --- | --- |
| **Cty** | **CtyCode** | **CtyName** | **CtySize** |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **State** | **StateCode** | **StateName** | **StateRgn** | **StateSize** | **StatePop** |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Con** | **ConID** | **ConName** | **ConPop** | **ConSize** |

With this in mind, there are transitive and derived dependencies that are present in third normal form.

* In the Person table, the following attribute that can be taken off is the PerAge.
  + Deleting the PerAge is transitive to PerDofB since it is dependent on PerDofB, but if we were to remove PerDofB, it will be difficult to know when exactly the PerAge will happen since many people can have the same age.
* In the Cty table, there are no dependencies that are present so then it is good.
* In the State table, the following attribute that can be removed is StateSize.
  + Deleting the StateSize is derived to CtySize since it is dependent on CtySize because the CtySize is made up of the StateSize and can be added in total to get the size of the state. If it were to be opposite, deleting the CtySize will cause the StateSize to guess the length of how large the city is that will be within that state.
* In the Con table, there will be two attributes delete (ConPop and ConSize.)
  + Deleting the ConSize is derived to the CtySize since StateSize has been removed. ConSize is dependent to all the cities that are inside of the country and can be calculated.
  + Similar to ConSize, deleting ConPop is derived to StatePop that has the following population within the state. It can be easily calculated when adding all the population of the total states inside the country to get the ConPop.

**Third Normal Form:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Person** | **PerSSN** | **PerName** | **PerDofB** | **PerAdd** |

|  |  |  |  |
| --- | --- | --- | --- |
| **Cty** | **CtyCode** | **CtyName** | **CtySize** |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **State** | **StateCode** | **StateName** | **StateRgn** | **StatePop** |

|  |  |  |
| --- | --- | --- |
| **Con** | **ConID** | **ConName** |

**Country(ConID, ConName)**

**State(StateCode, StateName, StateRgn, StatePop, ConID\*)**

**City(CtyCode, CtyName, CtySize, StateCode\*)**

**Person(PerSSN, PerName, PerDofB, PerAdd, CtyCode\*)**

1. **Draw your ERD based on fully normalized table (Reverse Engineering).**

**Diagram

Description automatically generated**