

[CS 315] Principles of Database Management : Assignment 1

Aniket Sanghi (170110)
sanghi@iitk.ac.in

Due: Feb 14, 2020

1 Relation Schemas

1. **states** (StateCode, DistrictCode, AreaName)
Since the above information was the same in each table (redundant) so better keep it at a single place. Referred to the state via just the state code in other tables.
2. **multilingual_education** (State Code, Area Type, Education, Bilingual Persons, ...)
I kept this table in almost the same form, just removed the redundant information of DistrictCode etc.
3. **multilingual_age** (State Code, ... Agemin, Agemax, ...)
Removed the extra information as above. Added 2 new columns of **Agemin** and **Agemax** that stores the minimum and maximum age corresponding to each group. This helped in 2 of the queries as they required mathematical operations on the age.
4. **population** (State Code, ... , Population_Persons...)
This table contains the total population, literate population and illiterate population. Added the same 2 extra columns here too to make use of integer properties of age in the queries.
5. **literates** (State Code, , Middle, Metric,)
This table contains all the given details of population only for the literate section divided by the different categories of literate people

2 INSERT INTO Database

- Bash script "insert_into_database.sh" make schemas for the population.db and insert data from csv file into it by manipulating it accordingly.
- Mostly all bash queries I have written are in pipelined form where they are manipulating the data in the flow, sed is replacing, cut is finding fields etc.

3 QUERY logic

Following is the quick logic behind each query

1. Find the rows with data of the total of all ages of all area types and then compute $\text{trilingual}/\text{total} \times 100$ order by state
2. Join the 2 tables with different type of age groups, then using $\text{agemin}/\text{agemax}$ keep only those tuples which have an intersection. Then finally group them according to trilingual table Age Groups and finally compute max multilingual percentage age group
3. Take union of 2 tables, one with $\text{no.of.males}/\text{no.of.females}$ and other with $\text{no.of.females}/\text{no.of.males}$ and then find the max
4. Directly join the tables and find the total row, then subtract to get the answer
5. Using the agemin and agemax in each row compute the average population of each state and then take max

4 How To Run

Creating the Database

- `chmod 700 insert_into_database.sh`
- `./insert_into_database.sh`

The above execution requires all the 3 csv files to be in this folder. Hence I have submitted assignment with the csv files in it. It then creates the database in `population.db`

Running Queries

- `chmod 700 query.sh`
- `./query.sh QUERY_NUMBER`

In place of `QUERY_NUMBER` add the corresponding query like the following for query 2
`>>> ./query.sh 2`