

Report for Data Models and Query Language Project

1st Arnav Ashank

School of Engineering and Applied Sciences
University at Buffalo
Buffalo, United States
arnavash@buffalo.edu

2nd Mariya Johar

School of Engineering and Applied Sciences
University at Buffalo
Buffalo, United States
mariyajo@buffalo.edu

3rd Nishi Ajay Mandhana

School of Engineering and Applied Sciences
University at Buffalo
Buffalo, United States
nishiaja@buffalo.edu

Abstract—Census Database of citizens enables the government authorities to have statistical information about the citizens of the country which helps them in planning for different economic and social reforms. In this project we designed a database for census data. It gives a complete demographic overview of the country.

Index Terms—census data, Postgres, State, District, Population

I. INTRODUCTION

Census data is collected every ten years by government officials. The Database of citizens enables the government authorities to have statistical information about the citizens of the country which helps them in planning for different economic and social reforms. For example the literacy data gives information about education level of the citizens in a region, if a region has lower literacy rate then laying out policies particularly for that region will not only be helpful for the citizens but also help in proper allocation of government funds, if the same policy will be implemented on a national level then will lead to more expenditure on the revenue and the results will not be remarkable for those regions that already have reached saturation in literacy rates. Similarly for utilities data in each region will give a clear picture of which region is more developed. It will be useful for the government in efficient planning and allocation of revenue for growth and development of the country.

II. TARGET USER

The main target user is Government official. It helps the government to give a clear picture of social and economic status of a region and efficient allocation of the revenue for the growth and development of the citizens and the country. Not only from the growth and development perspective, also in case of a calamity it will help the government assess the prior

condition and how much time and revenue will be required to bring it back to normal. This database can be updated at regular intervals which will help the government to keep track of the implication of their policies, its implementation and impact on the society.

III. DATASET

A. Schema

```
State (  
  State_code INTEGER PRIMARY KEY,  
  State_Name VARCHAR(100),  
  No_of_District INTEGER);
```

```
District (  
  District_code INTEGER PRIMARY KEY,  
  State_code INTEGER,  
  District_Name VARCHAR(100),  
  Rural_household bigint,  
  Urban_household bigint,  
  Foreign Key (State_code)  
  references State(State_code)  
  on delete cascade)
```

```
Literacy (  
  Sr no. INTEGER PRIMARY KEY,  
  District_code bigint,  
  State_code bigint,  
  Male_Literate bigint,  
  Female_Literate bigint,  
  Below_Primary_Education bigint,  
  Primary_Education bigint,
```

```

Middle_Education bigint,
Secondary_Education bigint,
Higher_Education bigint,
Graduate_Education bigint,
Other_Education bigint,
Foreign Key (State_code)
references State(State_code)
on delete cascade,
Foreign Key (District_code)
references District(District_code)
on delete cascade)

```

```

Profession (
Srno. INTEGER PRIMARY KEY,
District_code INTEGER,
State_code INTEGER,
Frontline_Workers bigint,
Industrial_Workers bigint,
Non_Workers bigint,
Cultivator_Workers bigint,
Agricultural_Workers bigint,
Household_Workers bigint,
Other_Workers bigint,
Foreign Key (State_code)
references State(State_code)
on delete cascade,
Foreign Key (District_code)
references District(District_code)
on delete cascade)

```

```

Population (
Srno. INTEGER PRIMARY KEY,
District_code INTEGER,
State_code INTEGER,
Total_Population bigint,
Male bigint,
Female bigint,
Age_Group_0_29 bigint,
Age_Group_30_49 bigint,
Age_Group_50 bigint,
Age_not_stated bigint,
Foreign Key (State_code)
references State(State_code)
on delete cascade,
Foreign Key (District_code)
references District(District_code)
on delete cascade)

```

```

Utilities (
Srno. INTEGER PRIMARY KEY,
District_code INTEGER,
State_code INTEGER,
LPG_or_PNG_Households bigint,

```

```

Housholds_with_Electric_Lighting bigint,
Households_with_Internet bigint,
Households_with_Computer bigint,
Foreign Key (State_code)
references State(State_code)
on delete cascade,
Foreign Key (District_code)
references District(District_code)
on delete cascade)

```

The State table has state code as the primary key, the district table has district code as the primary key. Each of the tables Literacy, Profession, Population, Utilities have Srno as their primary key which is distinct for each of them and is independent of the data in the database, and they have state code and district code as foreign key referred from state and district tables respectively. On delete cascade is the foreign key constraint used in the database for each table.

B. Dataset

1) State:

- State code- It is a code assigned to a state. It is defined as an integer type and is the primary key for the state table.
- State name- Name of the state. It is defined as a string type.
- No of district- Number of districts in a given state. It is defined as an integer type.

2) District:

- District code- It is a code assigned to a district of a given state. It is defined as an integer type and is the primary key for the district table.
- State code- It is a code assigned to a state. It is defined as an integer type and is foreign key for the district table. It is defined as an integer type.
- District Name- Name of the district. It is defined as a string type.
- Rural household- It is the number of rural households in the given district of a state. It is of bigint type.
- Urban household- It is the number of urban households in the given district of a state. It is of bigint type.

3) Literacy:

- District code- It is a code assigned to a district of a given state. It is defined as an integer type and is foreign key for the literacy table.
- State code- It is a code assigned to a state. It is defined as an integer type and is foreign key for the literacy table.
- Male Literate- It is defined as the number of literate males in the given district of a state. It is of bigint type.
- Female Literate- It is defined as the number of literate females in the given district of a state. It is of bigint type.
- Below Primary Education- It is the population in the given district of a state with below primary level of education. It is of bigint type.

- Primary Education- It is the population in the given district of a state with below primary level of education.It is of bigint type.
- Middle Education-It is the population in the given district of a state having a middle level of education.It is of bigint type.
- Secondary Education- It is the population in the given district of a state with secondary level of education.It is of bigint type.
- Higher Education- It is the population in the given district of a state with higher level of education.It is of bigint type.
- Graduate Education- It is the population in the given district of a state having completed graduate level of education.It is of bigint type.
- Other Education- It is the population in the given district of a state belonging to other education.It is of bigint type.

4) Profession:

- District code-It is a code assigned to a district of a given state. It is defined as an integer type and is foreign key for the profession table.
- State code-It is a code assigned to a state. It is defined as an integer type and is foreign key for the profession table.
- Frontline Workers- It includes the population involved in frontline jobs such army, navy, air force, medical etc. It is of bigint type.
- Industrial Workers- It includes the population involved in any kind of industrial job such IT, Iron and Steel Industry etc. It is of bigint type.
- Non Workers- It includes the population not involved in any kind of job for at least an year. It is of bigint type.
- Cultivator Workers- It includes the population engaged in cultivation of land owned or held from Government or held from private persons or institutions for payment in money, kind or share involved in any kind of cultivating job for at least an year. It is of bigint type.
- Agricultural Workers It includes the population who works on another person's land for wages in money or kind or share. The person does own the land on which he/she works. It is of bigint type.
- Household Workers- Household workers are defined as those where one or more members of the household are at home or within the village in rural areas and only within the precincts of the house where the household lives in urban areas. It is of bigint type.
- Other Workers- Workers other than cultivators, agricultural workers, Household Workers, are defined as 'Other Workers' (OW). Examples of such types of workers are government servants, municipal employees, teachers, factory workers, plantation workers, those engaged in trade, commerce, business, transport, banking, mining, construction, political or social work, priests, entertainment artists, etc. It is of bigint type.

5) Population:

- District code-It is a code assigned to a district of a given state. It is defined as an integer type and is foreign key for the population table.
- State code-It is a code assigned to a state. It is defined as an integer type and is foreign key for the population table.
- Total Population- It is the total population in the given district of a state. It is of bigint type.
- Male- It is the total males in the given district of the state. It is of bigint type.
- Female- It is the total females in the given district of the state. It is of bigint type.
- Age Group 0 29- It is the total population in the 0 to 29 age group. It is of bigint type.
- Age Group 30 49- It is the total population in the 30 to 49 age group. It is of bigint type.
- Age Group 50- It is the total population having age 50 and above. It is of bigint type.
- Age not stated- It is the total population whose age was not stated while census data collection. It is of bigint type.

6) Utilities:

- District code-It is a code assigned to a district of a given state. It is defined as an integer type and is foreign key for the utilities table.
- State code-It is a code assigned to a state. It is defined as an integer type and is foreign key for the utilities table.
- LPG or PNG Households- It is the count of households in the given district of the state having LPG or PNG supply. It is of bigint type.
- Housholds with Electric Lighting- It is the count of households in the given district of the state having Electric lightning. It is of bigint type.
- Households with Internet- It is the count of households in the given district of the state having internet connectivity. It is of bigint type.
- Households with Computer- It is the count of households in the given district of the state having computers. It is of bigint type.

C. Data modification and departure from milestone1

IV. WEBSITE

A. Backend

1) *Setting up python and database connection:* Step 1- Importing data into a raw database from csv file We have imported csv data and loaded it into 'data' table in the database. We use this 'data' table as the master table to insert the data into the rest of the tables: state, district, profession, population, utilities, literacy. We connect the database using the python API's and use those to push the data into the database.

B. Frontend

1) *Login page:* The login page is used to login into the application, using the data fields from login table of the database.

2) *Main Page*: The Main Page has several tab contents to divert to the several queries that we have written.

V. SOME INTERESTING INTERPRETATION IN OUR PROJECT

Q1 What is the percentage growth of state given internet and computer as amenities print result in descending order of percentage growth and sum of internet and computer in the state?

```
WITH ctel
AS (SELECT state_code,
i,c,( i + c ) AS sum_u
FROM (SELECT state_code,
Sum(households_with_internet) AS i,
Sum(households_with_computer) AS c
FROM utilities
GROUP BY ( state_code )
ORDER BY state_code)t)

SELECT s.state_name,
sum_u AS sum_internet_computer,
Round(sum_u * 100 / t.s) AS total_growth_percentage
FROM ctel c
CROSS JOIN (SELECT Sum(sum_u) AS s
FROM ctel) t
JOIN state s
ON c.state_code = s.state_code
ORDER BY sum_internet_computer DESC,
total_growth_percentage DESC
```

Output:

image need to be inserted

Q2 What is total population of male and female in country?

```
SELECT Sum(total_male_in_district)
AS total_male_in_country,
Sum(total_female_in_district)
AS total_female_in_country
FROM (SELECT d.district_code,
Sum(male) AS total_male_in_district,
Sum(female) AS total_female_in_district
FROM district d
JOIN population po
ON d.district_code = po.district_code
GROUP BY d.district_code
ORDER BY d.district_code)t
```

Output: image need to be inserted

Q3 What is state wise number of total agricultural workers and total Cultivator workers in descending order of sum of all the agricultural and cultivator ?

```
SELECT s.state_name,
Sum(agricultural_workers)
AS total_agriculture_worker_in_state,
Sum(cultivator_workers)
```

```
AS total_cultivator_workers
FROM profession ps
JOIN state s
N ps.state_code = s.state_code
GROUP BY s.state_name
ORDER BY total_cultivator_workers DESC,
total_agriculture_worker_in_state DESC
```

Output: image need to be inserted

VI. QUERY OPTIMIZATION

1) Procedure Function:

```
create function
get_male_count(user_state_code int)
returns int
language plpgsql
as
$male_count$
declare male_count integer;
begin
select sum(male)
into male_count
from population
where state_code = user_state_code;
return male_count;
end;
$male_count$;

select get_male_count(10);
select sum(male) from population
where state_code = 6;
```

2) Using Indexing on State code:

```
CREATE INDEX idx_state_code
ON population(state_code);

CREATE INDEX idx_state_code_utilities
ON utilities(state_code);
```

3) Using Common Table:

```
WITH ctel
AS (SELECT state_code,
i,c,( i + c ) AS sum_u
FROM (SELECT state_code,
Sum(households_with_internet) AS i,
Sum(households_with_computer) AS c
FROM utilities
GROUP BY ( state_code )
ORDER BY state_code)t)
SELECT s.state_name,
sum_u AS sum_internet_computer,
Round(sum_u * 100 / t.s)
AS total_growth_percentage
```

```
FROM   ctel c
CROSS JOIN (SELECT Sum(sum_u) AS s
FROM     ctel) t
JOIN state s
ON c.state_code = s.state_code
ORDER  BY sum_internet_computer DESC,
total_growth_percentage DESC
```

VII. REFERENCES