

Nicholas Nikas  
260870980

## COMP 189: Homework #4

Assigned March 15, 2021

Due March 26, 2021

64 points total

**Instructions:** For each problem, show all your work (required for credit). For answers requiring written answers, while no more than five or six sentences are expected, sufficient justification must be given for any position, opinion, or perspective taken.

**Submission Instructions:** submit your solutions in PDF format through MyCourses Assignments.

### Technical Exercises

#### 1. Database schema (6 points)

Reconstruct the schema for all tables in the World database. Indicate foreign keys and primary key using notation as described in lecture.

#### City

ID: integer (Primary Key)

Name: text

CountryCode: text (Foreign Key to Country)

District: text

Population: integer

#### Country

Code: text (Primary Key)

Name: text

Continent: text

Region: text

SurfaceArea: float

IndepYear: integer

Population: integer

LifeExpectancy: float

GNP: float

GNPOld: float

LocalName: text

GovernmentForm: text

HeadOfState: text

Capital: integer (Foreign Key to City)

Code2: text

### Country Language

CountryCode: text (Primary Key)

Language: text

IsOfficial: text

Percentage: float

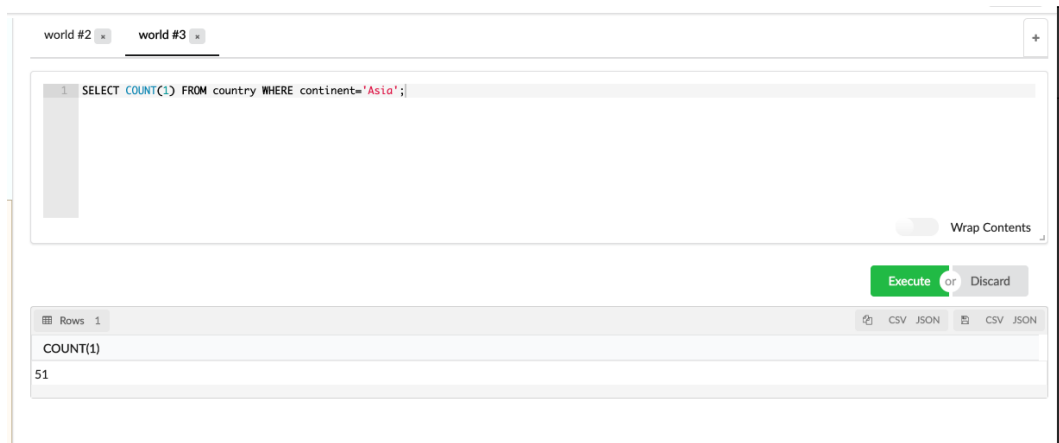
### 2. Basic Queries (15 pts)

Construct and run SQL queries to answer the following questions. In your answer show both the SQL query (exactly one for each question) you used and also the answer output by the database system – this can be a shrunk screenshot of the relevant details.

1. How many countries are in Asia?

```
SELECT COUNT(1) FROM country WHERE continent='Asia';
```

51 countries



2. How many countries are in Asia with fewer than 5 million people?

```
SELECT COUNT(1) FROM country WHERE continent='Asia' AND Population < 5000000;
```

19 countries

The screenshot shows a SQL query editor with a single query: `1 SELECT COUNT(1) FROM country WHERE continent='Asia' AND Population < 50000000;`. The query is highlighted in yellow. Below the editor, there are buttons for 'Execute' (green) and 'Discard' (grey). The results pane shows a table with one row: 

COUNT(1)
19

. The interface also includes a 'Wrap Contents' toggle and export options for CSV and JSON.

3. How many countries have Elisabeth II as their head of state?

`SELECT COUNT(1) FROM country WHERE HeadOfState='Elisabeth II';`  
35 countries

The screenshot shows a SQL query editor with a single query: `1 SELECT COUNT(1) FROM country WHERE HeadOfState='Elisabeth II';`. The query is highlighted in yellow. Below the editor, there are buttons for 'Execute' (green) and 'Discard' (grey). The results pane shows a table with one row: 

COUNT(1)
35

. The interface also includes a 'Wrap Contents' toggle and export options for CSV and JSON.

4. How many Government Forms are there in the database?

`SELECT DISTINCT GovernmentForm FROM country;`  
35 rows hence government forms

world #2

world #3

+

1

SELECT DISTINCT GovernmentForm FROM country;

Wrap Contents

Execute

or

Discard

Rows 35

CSV JSON

CSV JSON

GovernmentForm
Nonmetropolitan Territory of The Netherlands
Islamic Emirate
Republic
Dependent Territory of the UK
Parliamentary Coprincipality
Emirate Federation
Federal Republic
US Territory

5. What are the 5 least populous countries in South America? (your query must return exactly the 5 countries in your answer).

SELECT Name FROM country WHERE continent='South America' ORDER BY Population ASC LIMIT 5;  
Falkland Islands, French Guiana, Suriname, Guyana, Uruguay

world #2

world #3

+

1

SELECT Name FROM country WHERE continent='South America' ORDER BY Population ASC LIMIT 5;

Wrap Contents

Execute

or

Discard

Rows 5

CSV JSON

CSV JSON

Name
Falkland Islands
French Guiana
Suriname
Guyana
Uruguay

### 3. Messy data (10 pts)

The GovernmentForm field in the country database contains messy data – meaning that there are errors in it that really should have been corrected before creation of the database.

1. Identify two such errors.

One error is that it is very redundant where multiple countries have the same GovernmentForm types. Another error is that many GovernmentForm types have the same form with a small variation like Republic and Federal Republic for example.

2. How many countries are implicated by these errors?

For example, 122 countries are a Republic, 15 countries are a federal republic, 29 countries are a Constitutional Monarchy. So almost all countries are affected by this redundancy.

3. More generally, there really are a small number of possible Government Types – which could be better captured using a foreign key design pattern that we learned in class. Write out a new schema for the World database that employs foreign keys to reference GovernmentForms rather than a string field.

GovernmentForms

ID: integer (Primary Key)

Type: text

Country

Code: text (Primary Key)

GovernmentForm: integer (Foreign Key to GovernmentForms)

...rest

### 2. Advanced Queries (21 pts)

Construct and run SQL queries to answer the following questions. In your answer show both the SQL query (exactly one for each question) you used and also the answer output by the database system – this can be a shrunk screenshot of the relevant details. In every case, your query must return the exact value indicated.

1. What are the official languages spoken in countries in South America? (Each language should appear only once)

```
SELECT DISTINCT countrylanguage.Language, countrylanguage.IsOfficial
FROM countrylanguage JOIN country ON countrylanguage.CountryCode = country.Code
WHERE country.Continent = 'South America' AND countrylanguage.IsOfficial = 'T';
```

Spanish, Aymará, Ketšua, Portuguese, English, Guaraní

world #2 world #3

```

1 SELECT DISTINCT countrylanguage.Language, countrylanguage.IsOfficial
2 FROM countrylanguage JOIN country ON countrylanguage.CountryCode = country.Code
3 WHERE country.Continent = 'South America' AND countrylanguage.IsOfficial = 'T';

```

Wrap Contents

Execute or Discard

Rows 6

Language	IsOfficial
Spanish	T
Aimará	T
Ketsua	T
Portuguese	T
English	T
Guarani	T

- List the 8 countries with the most populous capital cities, in descending order by capital city population.

```

SELECT country.Name AS Country, city.name AS Capital, city.population AS CapitalPopulation
FROM country JOIN city ON city.ID = country.Capital ORDER BY city.population DESC LIMIT 8;

```

South Korea, Indonesia, Mexico, Russia, Japan, China, United Kingdom, Egypt

world #2

```

1 SELECT country.Name AS Country, city.name AS Capital, city.population AS CapitalPopulation
2 FROM country JOIN city ON city.ID = country.Capital ORDER BY city.population DESC LIMIT 8;

```

Wrap Contents

Execute or Discard

Rows 8

Country	Capital	CapitalPopulation
South Korea	Seoul	9981619
Indonesia	Jakarta	9604900
Mexico	Ciudad de México	8591309
Russian Federati...	Moscow	8389200
Japan	Tokyo	7980230
China	Peking	7472000
United Kingdom	London	7285000
Egypt	Cairo	6789479

- List the capital cities in South America that are in the 6 most populated countries.

```

SELECT city.name AS Capital, country.Name AS Country, country.population AS CountryPopulation
FROM country JOIN city ON city.ID = country.Capital
WHERE country.Continent = 'South America'
ORDER BY country.population DESC LIMIT 6;

```

Brasília, Bogotá, Buenos Aires, Lima, Caracas, Santiago de Chile

world #2

```

1 SELECT city.name AS Capital, country.Name AS Country, country.population AS CountryPopulation
2 FROM country JOIN city ON city.ID = country.Capital
3 WHERE country.Continent = 'South America'
4 ORDER BY country.population DESC LIMIT 6;
5

```

Wrap Contents

Execute or Discard

Rows 6

CSV JSON CSV JSON

Capital	Country	CountryPopulation
Brasília	Brazil	170115000
Santafé de Bogotá	Colombia	42321000
Buenos Aires	Argentina	37032000
Lima	Peru	25662000
Caracas	Venezuela	24170000
Santiago de Chile	Chile	15211000

#### 4. Give a breakdown of the number of countries for each form of government.

```
SELECT GovernmentForm, COUNT(GovernmentForm) FROM country GROUP BY GovernmentForm;
```

```

Administrated by the UN,1
Autonomous Area,1
Co-administrated,1
Commonwealth of the US,2
Constitutional Monarchy,29
Constitutional Monarchy (Emirate),1
"Constitutional Monarchy, Federation",4
Dependent Territory of Norway,2
Dependent Territory of the UK,12
Dependent Territory of the US,1
Emirate Federation,1
Federal Republic,15
Federation,1
Independent Church State,1
Islamic Emirate,1
Islamic Republic,2
Monarchy,5
Monarchy (Emirate),1
Monarchy (Sultanate),2
Nonmetropolitan Territory of France,4
Nonmetropolitan Territory of New Zealand,3
Nonmetropolitan Territory of The Netherlands,2
Occupied by Marocco,1
Overseas Department of France,4
Parlementary Monarchy,1
Parliamentary Coprincipality,1
Part of Denmark,2
People'sRepublic,1
Republic,122
Socialistic Republic,3
Socialistic State,1
Special Administrative Region of China,2
Territorial Collectivity of France,2
Territory of Australia,4
US Territory,3

```

world #2

```
1 SELECT GovernmentForm, COUNT(GovernmentForm) FROM country GROUP BY GovernmentForm;
```

Wrap Contents

Execute or Discard

Rows 35

CSV JSON CSV JSON

GovernmentForm	COUNT(GovernmentForm)
Administrated by the UN	1
Autonomous Area	1
Co-administrated	1
Commonwealth of the US	2
Constitutional Monarchy	29
Constitutional Monarchy (Emirate)	1
Constitutional Monarchy, Federa...	4
Dependent Territory of Norway	2
Dependent Territory of the UK	12
Dependent Territory of the US	1
Emirate Federation	1
Federal Republic	15
Federation	1
Independent Church State	1
Islamic Emirate	1

Log error Github

5. Give a breakdown of the number of South American countries for each form of government.

`SELECT GovernmentForm, COUNT(GovernmentForm) AS South_American_Countries  
FROM country WHERE continent='South America' GROUP BY GovernmentForm;`

Dependent Territory of the UK,1  
Federal Republic,3  
Overseas Department of France,1  
Republic,9

world #2

```
1 SELECT GovernmentForm, COUNT(GovernmentForm) FROM country WHERE continent='South America' GROUP BY GovernmentForm;
```

Wrap Contents

Execute or Discard

Rows 4

CSV JSON CSV JSON

GovernmentForm	COUNT(GovernmentForm)
Dependent Territory of the UK	1
Federal Republic	3
Overseas Department of France	1
Republic	9



6. What is the average life expectancy among Asian countries with a population less than 1 million?

```
SELECT AVG(LifeExpectancy) FROM country WHERE continent='Asia' AND Population < 1000000;  
69.35
```

The screenshot shows a SQL query editor with a tab labeled "world #2". The query entered is: `SELECT AVG(LifeExpectancy) FROM country WHERE continent='Asia' AND Population < 1000000;`. Below the query editor, there are buttons for "Execute" (in green) and "Discard". To the right of the "Execute" button is a toggle switch labeled "Wrap Contents". Below the query editor, a results table is displayed with the header "AVG(LifeExpectancy)" and a single row containing the value "69.35714". The table has a "Rows" header with the value "1". To the right of the table are buttons for "CSV" and "JSON".

7. What is the average life expectancy among Asian countries with a population greater than 1 million?

```
SELECT AVG(LifeExpectancy) FROM country WHERE continent='Asia' AND Population > 1000000;  
67.14
```

The screenshot shows a SQL query editor with a tab labeled "world #2". The query entered is: `SELECT AVG(LifeExpectancy) FROM country WHERE continent='Asia' AND Population > 1000000;`. Below the query editor, there are buttons for "Execute" (in green) and "Discard". To the right of the "Execute" button is a toggle switch labeled "Wrap Contents". Below the query editor, a results table is displayed with the header "AVG(LifeExpectancy)" and a single row containing the value "67.13636". The table has a "Rows" header with the value "1". To the right of the table are buttons for "CSV" and "JSON".

## Discussion

### 1. Security (12 pts)

A company stores all their passwords in a database, in a table called *User*. Using the webserver-database architecture we've discussed as well as the SQL language, explain how a database can make it easier for a hacker to steal *all* passwords belonging to users all at once. (Explaining the high-level attack they might use is not required, but might help provide a complete answer.)

The most common way the company's websites passwords can be stolen is by SQL injection. This occurs when the client asking for user input like a username and password and instead is given an SQL query that will unknowingly run on the company's database. When the client requests a username for example, you type 'UserName' and press enter what happens is the client sends an SQL query: `SELECT * FROM USERS WHERE UserName = 'UserName'` to the database. If the client does not have input validation, then you can write a smart input like `'21 OR 1=1'` which sends this query: `SELECT * FROM USERS WHERE UserName = 21 OR 1=1` which is a valid query since `1=1` is always true. If the Users table contains passwords, then the hacker can that easily retrieve all passwords in the company's database. So, a hacker can retrieve all passwords in a client-server database by simply typing `21 OR 1=1` in the user input. The best way to prevent SQL injection attacks is to have strong input validation in the client interface such as only accepting strings that match certain regular expressions (like an email must have an @ with a .com and thus won't accept `21 OR 1=1`) and rejecting wrong inputs before sending query to database. There are other possible vulnerabilities within a client-server-database software system where a hacker can bypass the user identification page and send malicious SQL queries to the database and steal all passwords hence it's important to ensure the database is well protected by attackers.