Explanation of Database

For my final project I created a database which keeps track of different information concerning United States Military Bases as well as the branches which operate and support them. I chose this data because I am interested in military history but I wanted to pick something modern so that I could get reliable data. This database would presumably be used for logistical reasons for military or infrastructure reasons. Each base is connected to its branch by a many to many relationship, meaning that individual branches could filter to see only bases pertinent to them. Branches would also be able to see important information regarding their manpower and funding. The funding of the branches is given in millions of dollars.

My data is sourced from these places:

https://www.statista.com/statistics/239290/budget-of-the-us-navy-and-the-us-marine-corps/

https://www.statista.com/statistics/239280/budget-plan-of-the-us-armed-forces-by-department-and-appropriation/

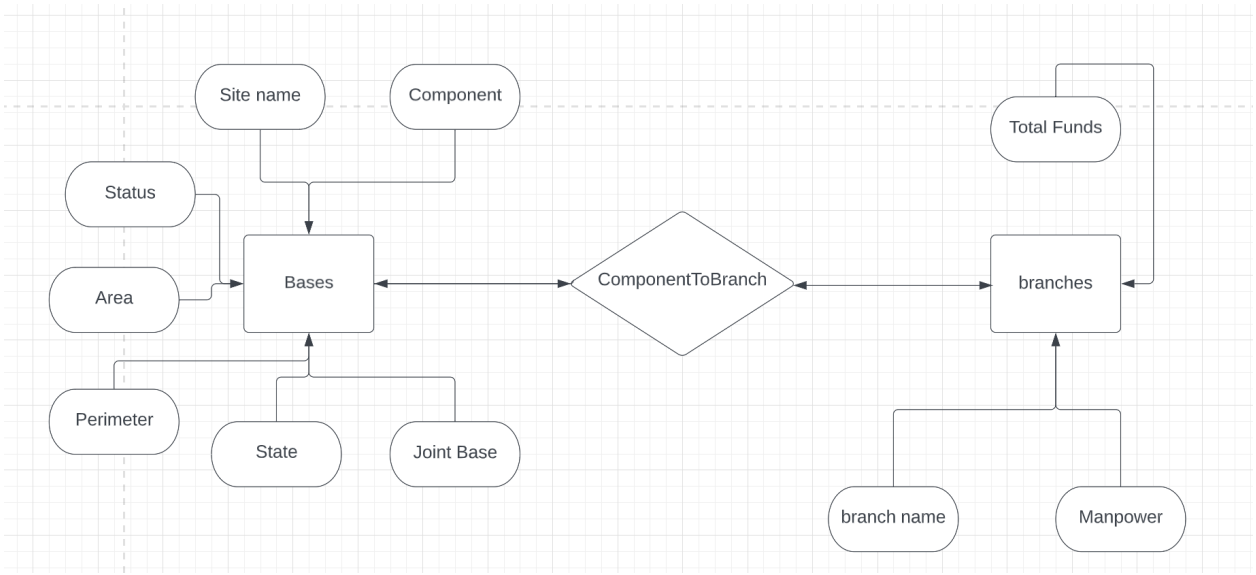
https://public.opendatasoft.com/explore/dataset/military-bases/table/

Relational Schema

Bases(<u>baseID</u>, component, site, joint base, state, status, perimeter, area)
Branches(<u>branch</u>, manpower, totalFunds,)
BranchToComponent(<u>component, branch</u>)

The database was designed such that the only functional dependencies occur between the key and the rest of the elements. I restructured and removed redundant data in order to make this happen. Because of this the database is automatically normalized. The bases database makes use of the auto incrementing baseID as its primary key as some sites share names. I have underlined the primary key, which functionally determines everything in each database.

ER Diagram



This diagram provides a high level view of the way the database is structured. It is designed to be human readable and easy to understand, so that any engineers who would have to work on this database could easily acquire a high level understanding of the way it is organized.

Questions

1. List all bases in Hawaii alongside which branch operates them
2. List the branches of military with active bases in California
3. Show the ten largest military bases by area
4. Show the ten largest bases by perimeter run by the airforce
5. Show all joint bases
6. List all branches in order of how many military bases they operate
7. List all States in order of how many military bases they contain
8. List the branches in order of how much funding they receive
9. Show all inactive bases and who used to operate them
10. Show all branches with more then 200,000 manpower

This questions are all answers with queries in the python file. I tried to pick a variety of questions to show that I understand more complicated operations such as inner joins, group by or order by to list a few.

Grading

I believe that the work I did for this project deserves an A. This project took me around 10 hours of total work to complete, most of it involving solving problems which I encountered along the way. The first problem I encountered was finding good data to use. It was difficult to find accurate data, especially regarding funding. The marine corp is a part of the navy, which means in order to separate their funding I needed to find a source on both and subtract the budget of the marines from the navy. Taking the raw data from the internet and deciding what I wanted to use and what not to use was also a struggle. I decided to remove some redundant columns, such as the row for Country which only ever could be United states or Guam (which was already clarified in the state column). Additionally, removing this row removed the functional dependency of state > country, meaning the data was now normalized. After I found all the data I needed and decided how to structure it, I started on the part of the project which took me the most time. Getting all of the raw data from the internet into my sql server was the most difficult part for me. There were too many rows of data to enter manually, and some edge cases continued to frustrate my efforts. I discovered that some military bases have the same site names so I decided to implement a baseID to use as a primary key. Overall I found this process to be both challenging and engaging. Finding raw data and turning it into a database which is friendly to work with is a skill that is important for any programmer who wishes to work on databases, and likely every programmer will have to interact with them in some way or another. After I got all the information imported and set up correctly I wrote the queries, which was the easiest part as I have the most practice doing it. In the end I think I learned a lot and I worked hard, so I hope you consider this when you give me my final grade. Thank you for your time.