Data modeling and pipelining

This section tests your data engineering capabilities. You can attempt this section or the next (system design) section of the quiz.  
  
Assume you are developing an application to bring out the latest in COVID related clinical trials. While there are any number of web sites providing dashboards on COVID related trials, none of them provide an easy way for users to get answers to questions of common interest such as:  
  
1. What are the leading candidates for vaccination/treatment as evidenced by:   
1. results obtained so far,   
2. number and nature of subjects enrolled in the trials,   
3. source of funding for the trials,   
4. experience of the biotech/pharma company that is the innovator,   
5. mechanism of action for the proposed drug/vaccine  
2. What are the limitations of leading vaccine/therapy candidates such as:  
1. Demography they’ve been tested on: E.g., age group, gender, race/nationality, social strata  
2. Ease of distribution and administration: E.g., number of visits to the clinic needed per patient, storage conditions needed to keep the drug potent and stable, devices needed for administering the medicine etc.  
3. Manufacturing capacity available to ramp-up production  
3. What are the co-morbidities that complicate COVID treatment and what treatment options are being tested in each case?  
  
In the questions below, you’ll be asked to design a data pipeline and a data model needed to create an application that will answer the above questions.   
  
Through this submission, you are expected to demonstrate your understanding of relational data modeling principles and your ability to grasp specifics of the domain you are modeling.   
  
To understand the basics of data modeling, we suggest using a formal text book (such as [https://opentextbc.ca/dbdesign01/](https://www.google.com/url?q=https://opentextbc.ca/dbdesign01/&sa=D&source=editors&ust=1622554599530000&usg=AFQjCNGRlOl6vL1p4PLDVaWtRCMN3W9PTQ)) or course (such as [https://www.coursera.org/learn/database-management/](https://www.google.com/url?q=https://www.coursera.org/learn/database-management/&sa=D&source=editors&ust=1622554599531000&usg=AFQjCNEYBHjYJ1GoNUeT_LFDdTp1lzjQFQ)).  
  
To understand the domain to model: Read the descriptions at [clinicaltrials.gov](https://www.google.com/url?q=http://clinicaltrials.gov&sa=D&source=editors&ust=1622554599531000&usg=AFQjCNHNLe-wdC5suKPZt6LJoncyEdAF_Q) for a few trials such as: [https://clinicaltrials.gov/ct2/show/NCT04794946](https://www.google.com/url?q=https://clinicaltrials.gov/ct2/show/NCT04794946&sa=D&source=editors&ust=1622554599531000&usg=AFQjCNH-JGnIFWVLidC6bkoUgDQOCbwnJw). Use the glossary at [https://clinicaltrials.gov/ct2/about-studies/glossary](https://www.google.com/url?q=https://clinicaltrials.gov/ct2/about-studies/glossary&sa=D&source=editors&ust=1622554599531000&usg=AFQjCNF6PQEMpLGipadeTnOS-M7aXes2AQ) for any clinical trials related terms you do not understand.

Question 1:

Using a diagramming/modeling tool such as the one at [https://dbdiagram.io](https://www.google.com/url?q=https://dbdiagram.io&sa=D&source=editors&ust=1622554599532000&usg=AFQjCNE1YIeGjg0E6VIQvQw-tzjGrru8fg), draw an entity-relationship (ER) diagram providing a data model for the proposed application. Provide here a link to your work.

-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

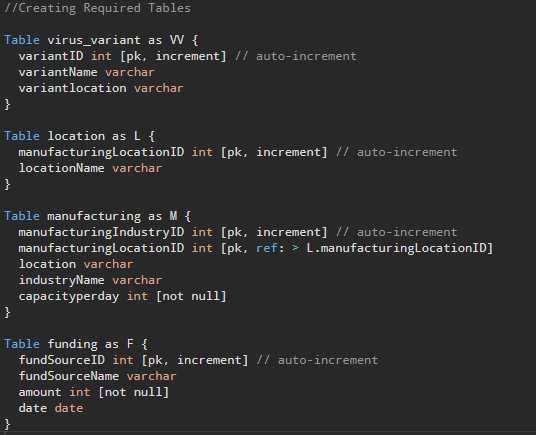
**ER Model/Diagram Link:**

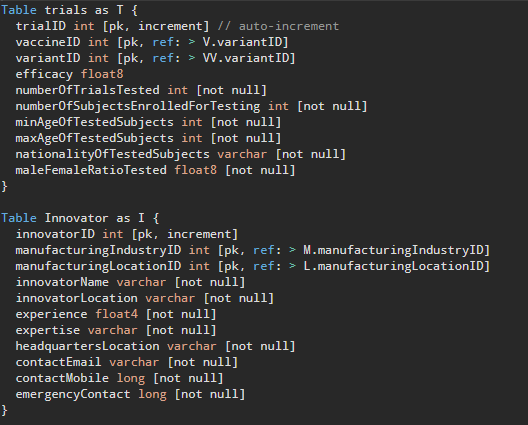
<https://dbdiagram.io/d/60b6523db29a09603d178023>

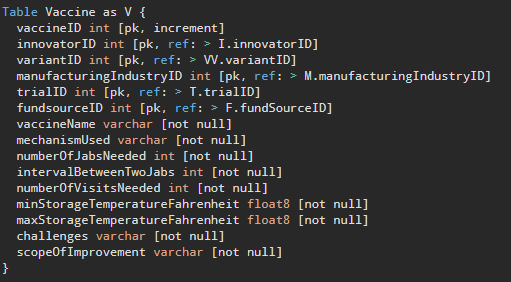
**Entity-Relationship model:**



**Script for ER model:**







**Brief Explanation:**

1. I have created 7 tables for this activity, namely: **vaccine**, **trials**, **innovator**, **manufacturing**, **virus\_variant**, **location** & **funding**.
2. I have made all the fields as **[not null]** as all the data here is extremely important and having sparse table is not at all recommended in the medical domain.
3. I have tried to follow all the fundamental rules of ER modelling and database creation.
4. I have tried to maintain normalization up to **3NF** as higher normalization levels may lead to loss of data.
5. This is just a first-hand implementation of the system and I can surely come up with a more detailed ER model if given more time.