Database Design For Mass Lotto Under a Microscope:

Database Technology: I have decided to use MySQL as EC2 under AWS using t2small storage choice. I will also be using a very small table to contain the needed data for dashboard, trending, and statistics called “MassCashTBL”. My route table ID created in AWS is “rtb-01b16da6956ba4a0b”. Below in Figure 1 is my table design.

The reason I am keeping the table simple is because I want to use the time to develop more dashboard and possibly event selections for the two main webpages. The dashboard, trending, and statistics are the crux of the web application. With the key pieces in place and utilizing a Crontab job scheduler, the idea to keep the parser pulling data from the Mass Cash Lottery (https://www.masslottery.com/games/lottery/search/results-lottery.html?game_id=12&mode=1&year=2012&month=12&x=5&y=13) into my table in MySQL. SQL statements combined with Javascript will be written to create various graphs and analytics on the webpages. My web application is focused on building a story for users based on available historical data.

Data Structure: From the design below, the main primary key is designated as the lottery ID. Each day when a lottery ticket is sold, an ID is created for that day, hence this is the main key which I will be using as both external and internal key. This will be stored as an integer. DrawDate is the date of each lottery instance. However, I am storing it as a varchar(20). It can be noted on the website above where data originates from; that the winning sequence should actually be stored as a string rather than tokenized via indexes, and stored separately like below design. Since this is preliminary, there will still be testing involved. If they’re stored separately, the parser script in Python will need to be revised include tokens. If it’s to be stored as is, then there should only be one field, WinningNum VARCHAR(20). In order to separate the numbers, MySQL will need to handle the substring functions to draw on the frequency by the order. The JackpotAmt field should always be “100000” and hence the data type is an INTEGER. The NumofWinners represent the number of people who won the lottery instance. The Store_Loc represent the towns where winning lottery instances were bought.

Roles of Each Data Element: In terms of the application, the most important data elements would be the five winning number fields (WinningNum1, WinningNum2, WinningNum3, WinningNum4, WinningNum5). From these, we can gauge the frequency, trending, and some statistics. Creating the various graphs and for example, Gaussian curve, we can use a distribution of the frequency in volume and DrawDate or Store_Loc. JackpotAmt doesn’t matter too much in this case, since Mass Cash winnings always equal to 100000. NumofWinners would be important as well because the frequency of winning numbers would have to factor in the number of winners, which would increase the mode of particular number counts in relation to other winning numbers.

(Figure 1-Table Design)