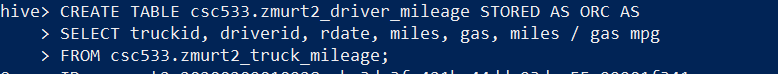
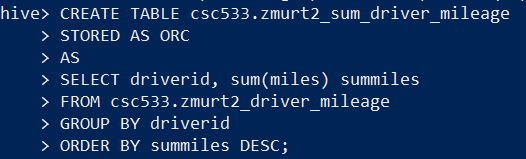
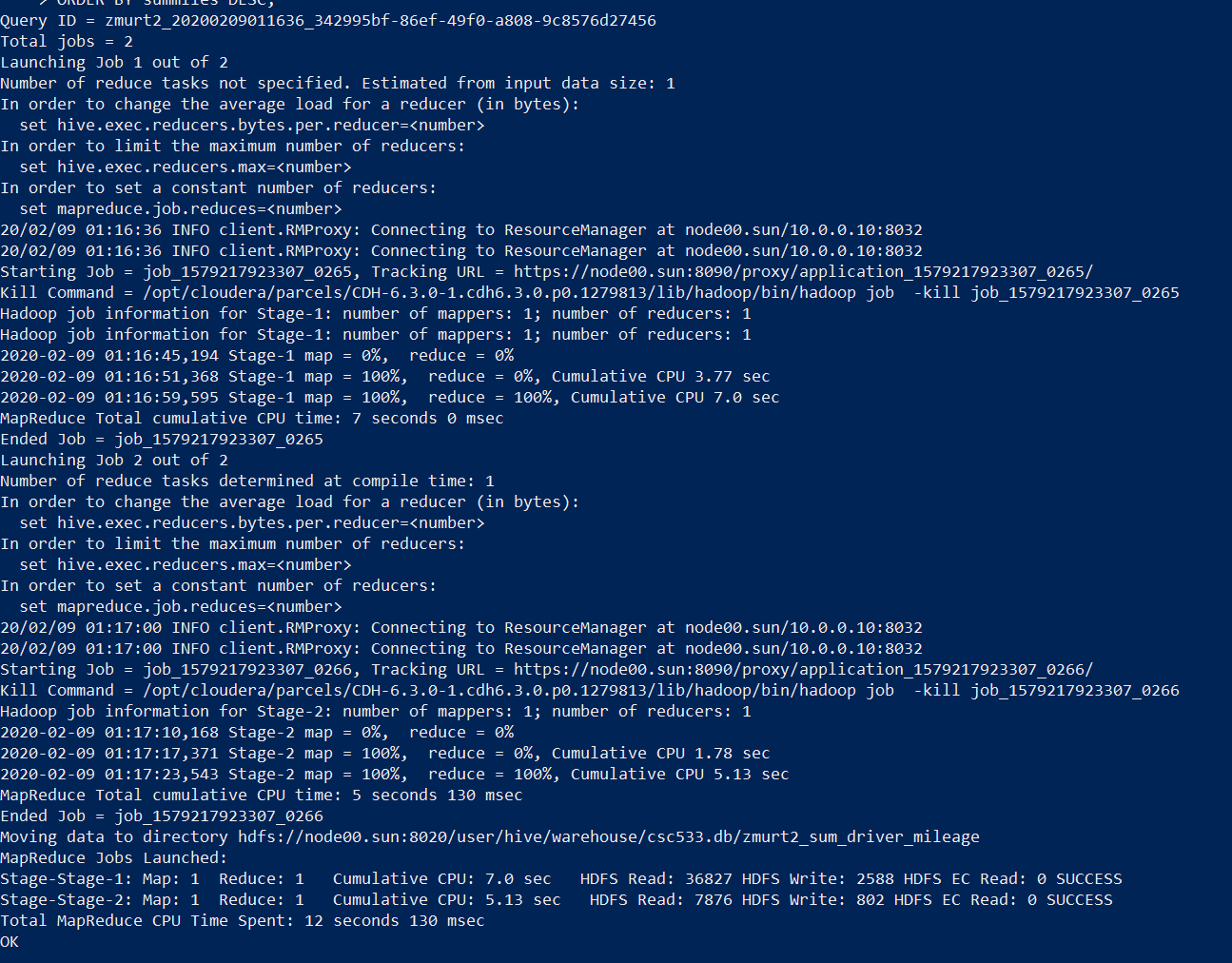
Used Hive to analyze derived data from the geolocation and trucks tables. The business objective is to better understand the risk the company is under from fatigue of drivers, over-used trucks, and the impact of various trucking events on risk. In order to accomplish this, we will apply a series of lateral transformations to the source data, mostly though SQL but stored in optimized row columnar formats to work well with Hive.

Question1: (Please Note: I’ve created two tables in this assignment, zmurt2\_driver\_mileage and zmurt2\_sum\_driver\_mileage. Both tables are stored as ORC file formatss. Apologies for the confusion!) zmurt2\_sum\_driver\_mileage is used for the aggregation functions and such.

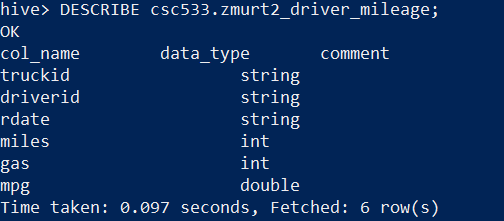




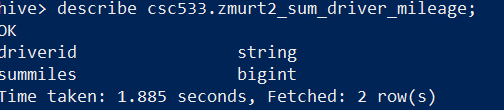


Question 2:

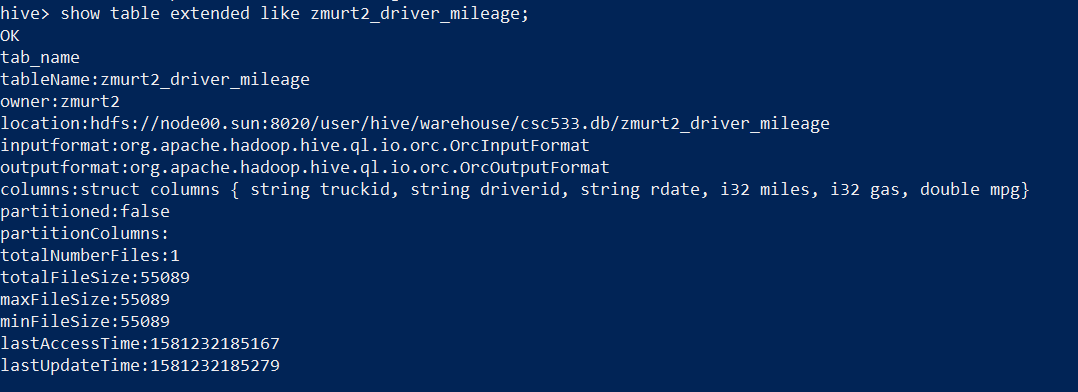
(i)



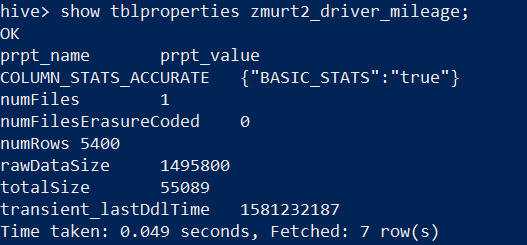
And my other table:

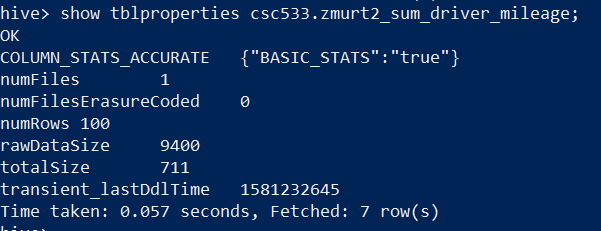


(ii)



Also…





Question 3:

