Aws big data project – writeup

Github link : https://github.com/manishsharma1402/AwsBigDataProject

First I have logged in aws console and created an IAM user with full access. Then I have created an s3 bucket and given all the required permissions. Further I have added my csv file in this s3 bucket. Then I moved to glue and selected crawler. then I have created a new crawler with create new crawler button. I have given a unique name to it. I have given the IAM user access to it then I have given source address and completed all the required steps. Then I have click run crawler button and waited till it runed successfully. Further I have clicked log button to go to see metadata catalog and found all the attributes with all its type. To convert csv to parquet I have clicked to new job and given it unique name. sourced placed was given s3 bucket and target was given as parquet file and completed required steps. Then I clicked on run to start the job after its successfully completion I have got the parquet files in s3 bucket. I have again repeated crawler step on parquet files. To get metadata logs of parquet files. Then I have moved to athena to run the query. I have typed the required query and successfully run it.