CREATE TABLE h1b1(s\_no int,case\_status string,  
employer\_name string, soc\_name string, job\_title string,  
full\_time\_position string,prevailing\_wage int,year string, worksite  
string, longitute double, latitute double )  
ROW FORMAT SERDE 'org.apache.hadoop.hive.serde2.OpenCSVSerde'  
WITH SERDEPROPERTIES (  
"separatorChar" = ",",  
"quoteChar" = "\""  
) STORED AS TEXTFILE;  
  
load data local inpath '/home/hduser/h1b.csv' overwrite into table  
h1b1;  
  
CREATE TABLE h1b2(s\_no int,case\_status string, employer\_name  
string, soc\_name string, job\_title string, full\_time\_position  
string,prevailing\_wage int,year string, worksite string, longitute  
double, latitute double )  
row format delimited  
fields terminated by '\t'  
STORED AS TEXTFILE;  
  
  
INSERT OVERWRITE TABLE h1b2 SELECT regexp\_replace(s\_no, "\t", ""),  
regexp\_replace(case\_status, "\t", ""), regexp\_replace(employer\_name,  
"\t", ""), regexp\_replace(soc\_name, "\t", ""),  
regexp\_replace(job\_title, "\t", ""),  
regexp\_replace(full\_time\_position, "\t", ""),  
regexp\_replace(prevailing\_wage, "\t", ""), regexp\_replace(year, "\t",  
""), regexp\_replace(worksite, "\t", ""), regexp\_replace(longitute,  
"\t", ""), regexp\_replace(latitute, "\t", "") FROM h1b1  
where case\_status != "NA";  
  
CREATE TABLE h1b(s\_no int,case\_status string, employer\_name  
string, soc\_name string, job\_title string, full\_time\_position  
string,prevailing\_wage int,year string, worksite string, longitute  
double, latitute double )  
row format delimited  
fields terminated by '\t'  
STORED AS TEXTFILE;  
  
INSERT OVERWRITE TABLE h1b SELECT s\_no,  
case when trim(case\_status) = "PENDING QUALITY AND COMPLIANCE REVIEW - UNASSIGNED" then "DENIED"  
when trim(case\_status) = "REJECTED" then "DENIED"  
when trim(case\_status) = "INVALIDATED" then "DENIED"  
else case\_status end,  
employer\_name, soc\_name, job\_title,  
full\_time\_position,prevailing\_wage,year, worksite, longitute, latitute  
FROM h1b2;