**Creating Flickr Hive table**

create external table flickr (fid string, latitude string, longitude string) row format delimited fields terminated by ',' location '/user/sds695/flickr\_data/';

**Creating Zillow Neighborhoods table**

create external table zillow1 (fid int, state string, county string, city string, name string, redionid string, polygons geometry) row format delimited fields terminated by ',' location '/user/sds695/Zillow\_data/';

**Increase the number of mappers and reducers and create functions to do the spatial join**

set hive.input.format=org.apache.hadoop.hive.ql.io.HiveInputFormat;

set mapred.map.tasks = 1000;

set mapreduce.job.reduces=100;

create temporary function st\_polygon as 'com.esri.hadoop.hive.ST\_Polygon';

create temporary function st\_point as 'com.esri.hadoop.hive.ST\_Point';

create temporary function st\_contains as 'com.esri.hadoop.hive.ST\_Contains';

**Performing spatial join on these two datasets**

CREATE TABLE flickr\_zillow AS SELECT zillow1.name, count(\*) cnt FROM zillow1 JOIN flickr WHERE st\_contains(st\_polygon(zillow1.geometry), st\_point(flickr.longitude, flickr.latitude)) GROUP BY zillow1.name ORDER BY cnt desc;

**Exporting the table to hdfs**

INSERT OVERWRITE DIRECTORY '/user/sds695/flickr\_zillow' select \* from flickr\_zillow;

**Exporting table from hdfs to dumbo(This command is run on dumbo not beeline)**

hdfs dfs -cat flickr\_zillow /000000\_0>flickr\_zillow.csv