Firstly, I scraped the photos (id, title and location) with the tag ‘graffiti’ in the NYC area from Flickr and saved it to a file. The code for the same can be found in the Scrape\_Flickr.py file attached with this.

Then, I transferred the file to Dumbo using scp command:

scp ./flickr\_scraped\_07\_21.txt dumbo:/home/srm699/RBDA/Project/

Then, from the Project folder on Dumbo, I transferred the file in HDFS using put command.

hdfs dfs -mkdir Project

hdfs dfs -put /home/srm699/RBDA/Project/flickr\_scraped\_07\_21.txt Project/