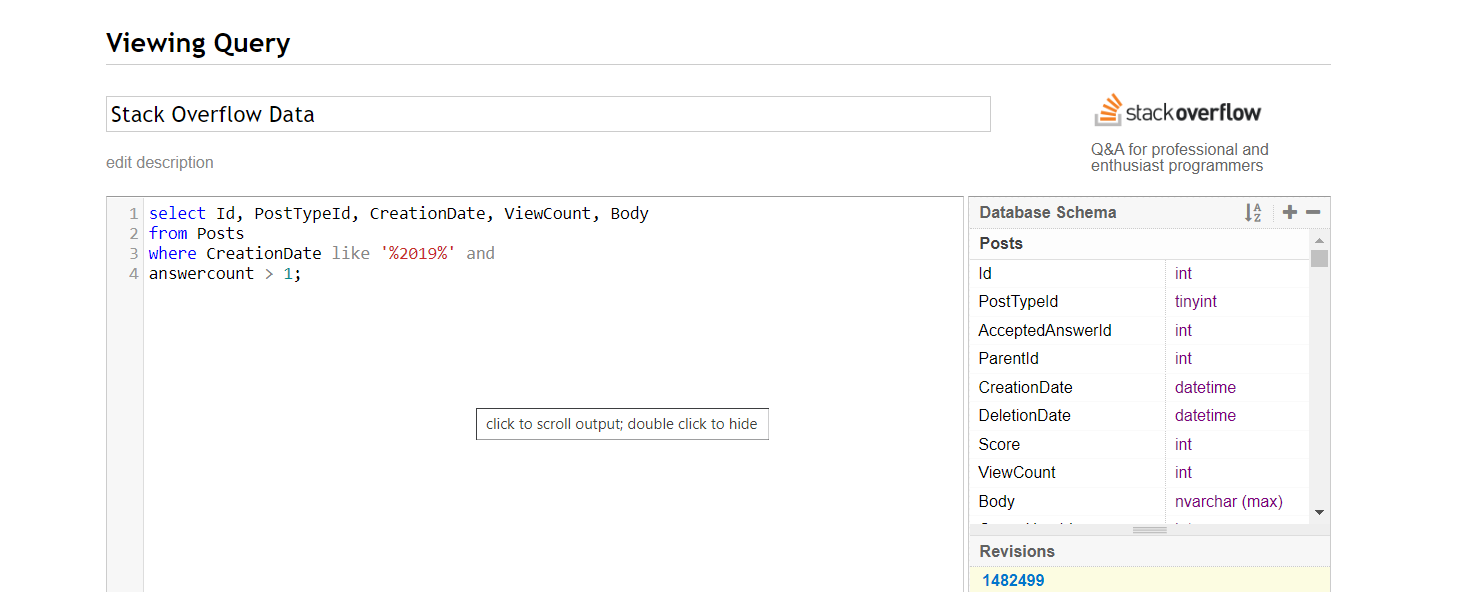
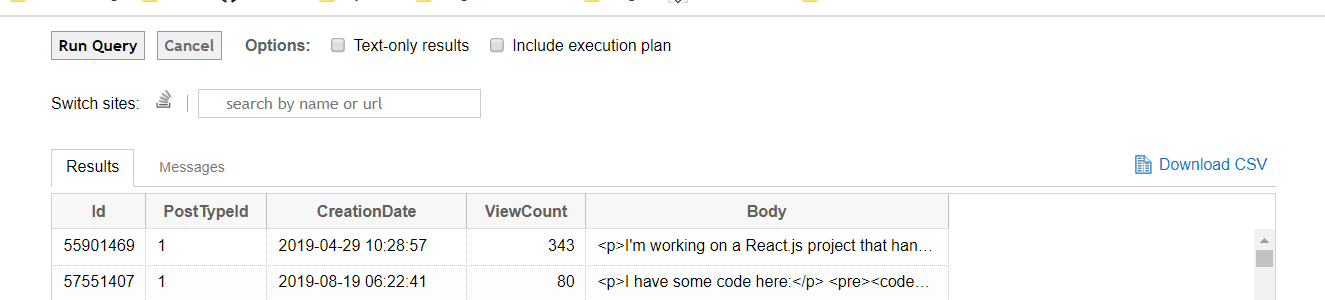
Inverted Index Project Weeks 7-8 Lab:

We leveraged data from Stack Overflow’s data exchange <https://data.stackexchange.com/stackoverflow/query/edit/1203802>, where I queried the Posts table obtaining posts from 2019 with answer counts, obtaining the ID, Post Type ID, View Count, and Body of the posts as shown below:



When I executed this query I was able to download the csv with the button/link to the right:



I now have the data required to load into HDFS, which will be the next step, this process is outlined in the PDF from Jupyter Notebook attached to this assignment.