**John Paul Garcia**

**Preetham Wilfred John**

**Big Data Analytics**

**Assignment Number: 1**

**Dr. Boetticher**

**Ms. Rekha Sampangiramaiah**

**Question 4:**

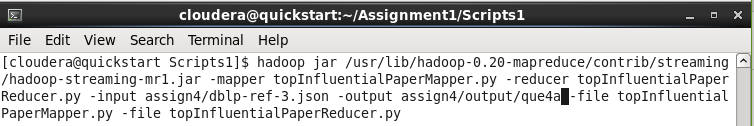
**Dataset: dblp-ref-3.json**

**Python Files:**

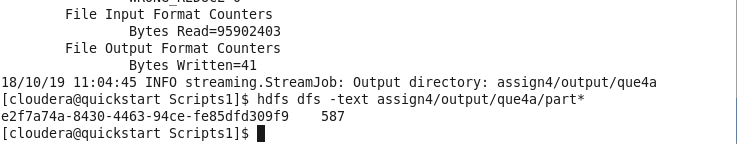
**topInfluentialPaperMapper.py**

**topInfluentialPaperReducer.py**

**Running the script through Hadoop:**



**Output:**



**Question 5:**

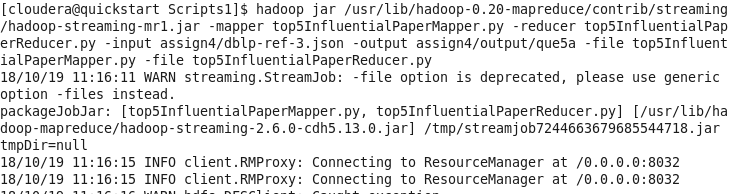
**Dataset: dblp-ref-3.json**

**Python Files:**

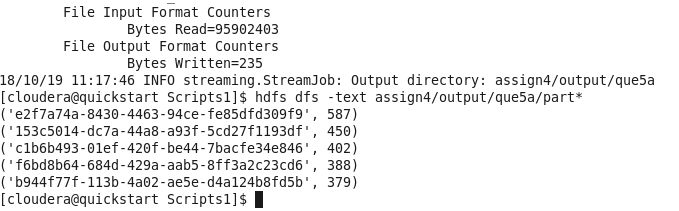
**top5InfluentialPaperMapper.py**

**top5InfluentialPaperReducer.py**

**Running the script through Hadoop:**



**Output:**



**Question 6:**

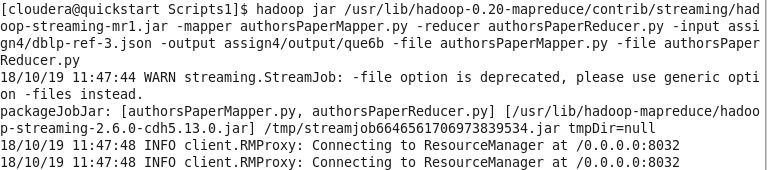
**Dataset: dblp-ref-3.json**

**Python Files:**

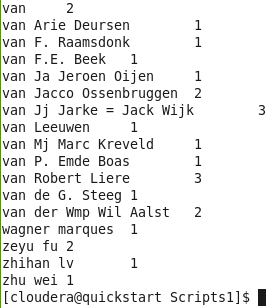
**authorsPaperMapper.py**

**authorsPaperReducer.py**

**Running the script through Hadoop:**



**Output:**



**Question 7:**

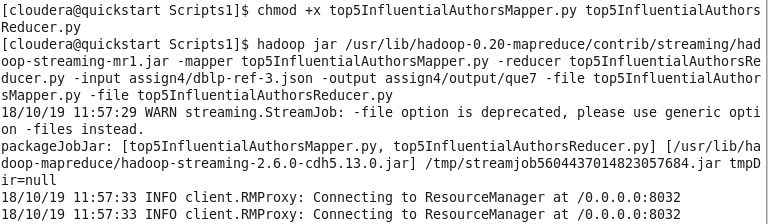
**Dataset: dblp-ref-3.json**

**Python Files:**

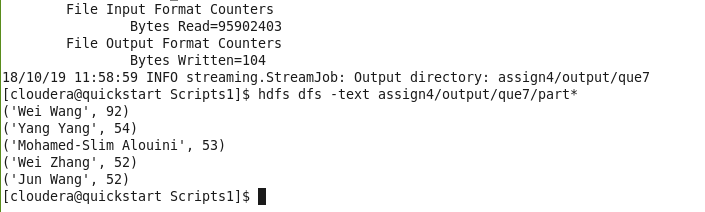
**top5InfluentialAuthorMapper.py**

**top5InfluentialAuthorReducer.py**

**Running the script through Hadoop:**



**Output:**



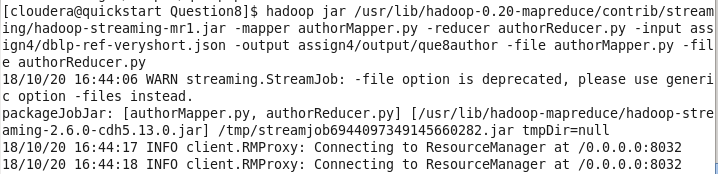
**Question 8:**

**Dataset: dblp-ref-veryshort.json**

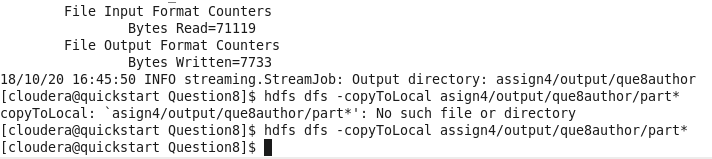
**Creating Inputs for the MapReduce by running the files below and downloading to local and uploading to HDFS under a single directory**

**authorMapper.py**

**authorReducer.py**

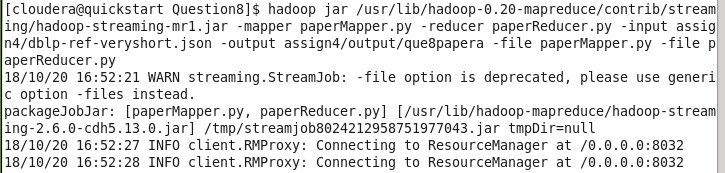


**Copying output:**

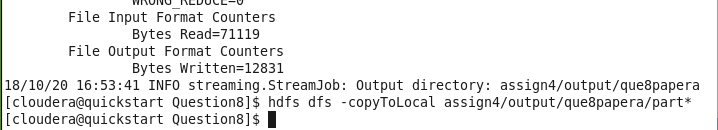


**paperMapper.py**

**paperReducer.py**



**Copying Output:**



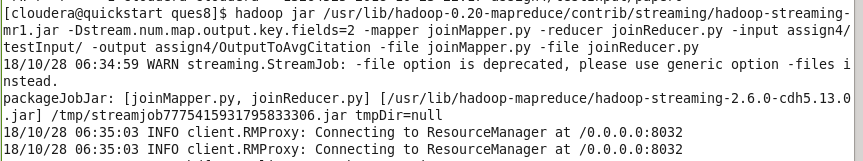
**Loading to HDFS:**

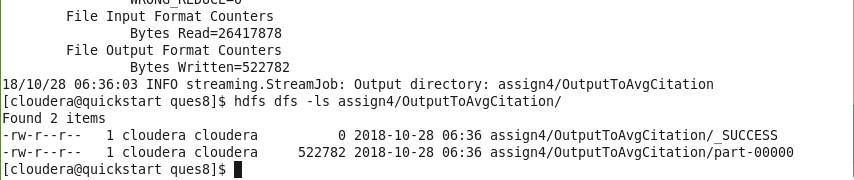


**Running the Join program to combine both the results generated in previous steps.**

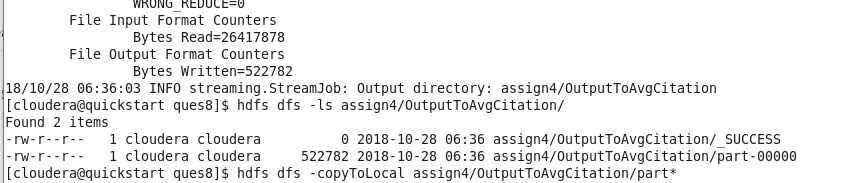
**joinMapper.py**

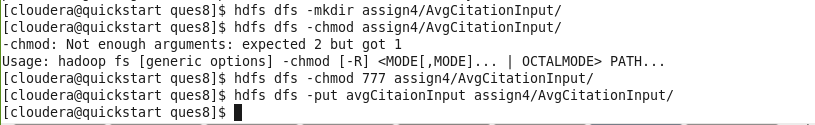
**joinReducer.py**





**Copying Output and loading to HDFS:**

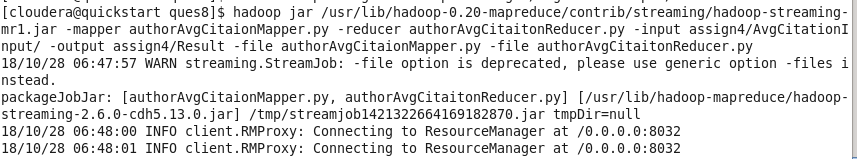




**Running the final programs by providing the input as the above path where the joined file is stored**

**authorAvgCitationMapper.py**

**authorAvgCitationReducer.py**



**Output:**

