Assignment 3

PROBLEM1-TASK3

Manraj Singh

B00877934

Problem-1

Task-3

- Algorithm for Map Reduce:

  - Read Data from each raw file [2].
  - Maintain a HashMap to count the keywords occurrences, one for each file.
  - In HashMap the key will be keyword name and value will be integer occurrence of that keyword.
  - Map all the occurrences of keywords in that file using indexOf and substring.
  - Count all the occurrences of the keywords in that file and store in its HashMap.
  - Store all the HashMaps generated for each raw file in a list.
  - Reduce the list of HashMaps.
  - Write the total count for each keyword across all the raw files in a json file: (./Output/MapReduceCount/MapREduceCount.json) as shown in *figure 1 [3]*

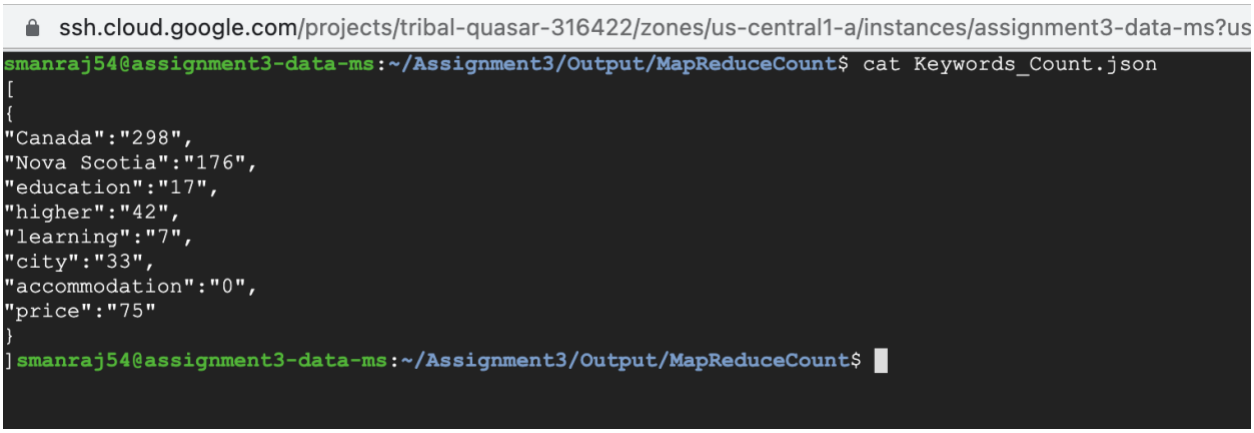

Figure 1: MapReduce Count of Keywords [1]

- Data in above figure clearly shows "*Canada*" keyword has highest count and "*accommodation*" keyword has the least count.

References

[1] Console.cloud.google.com. 2021. *Google Cloud Platform*. [online] Available at: <https://console.cloud.google.com/compute/instancesDetail/zones/us-central1-a/instances/assignment3-data-ms?project=tribal-quasar-316422&rif_reserved> [Accessed 5 July 2021].

[2] Newsapi.org. 2021. *News API – Search News and Blog Articles on the Web*. [online] Available at: <https://newsapi.org/> [Accessed 5 July 2021].