



EXPERIMENT NO: 04

Aim : Write a program to implement a word count program using Mapreduce

System Software Requirements : Java Compiler

Theory :

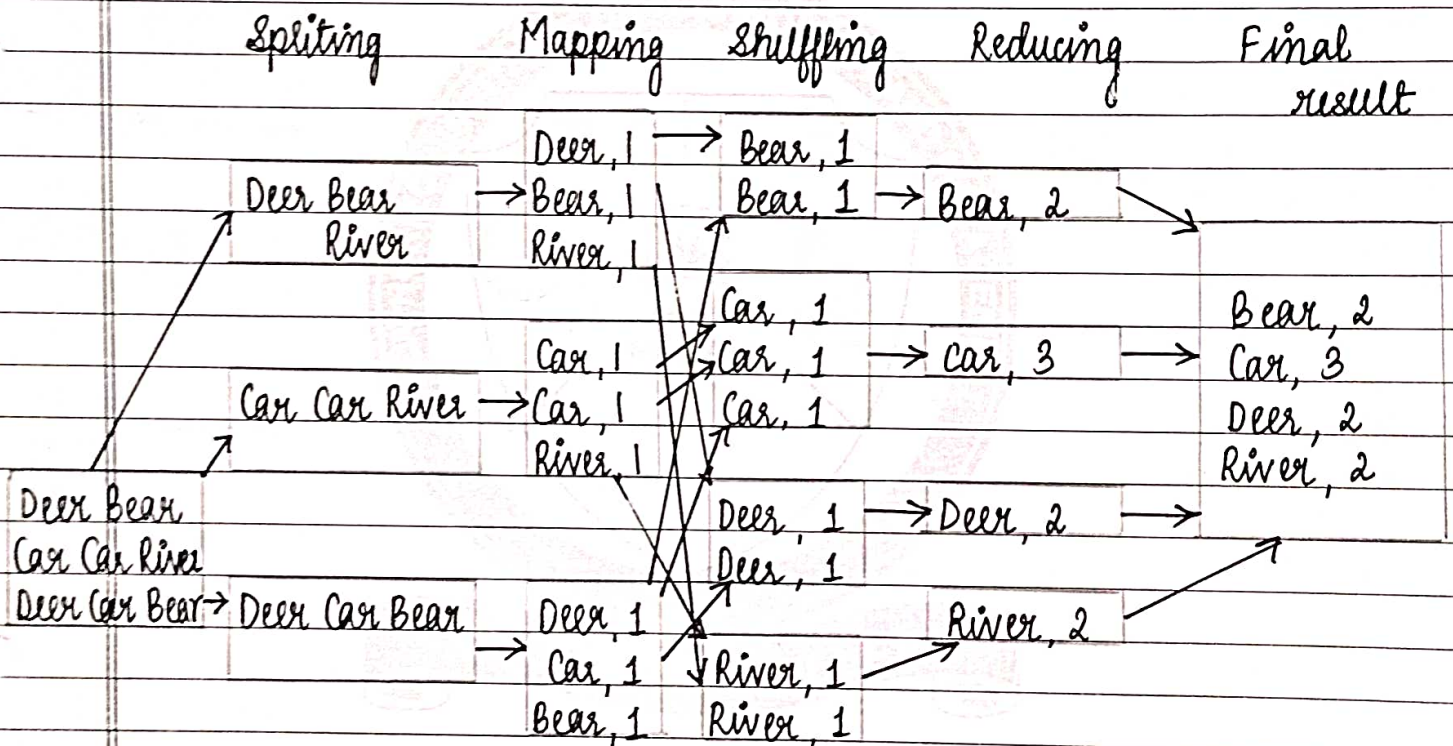
- * A MapReduce is a data processing tool which is used to process the data parallelly in a distributed form.
The mapreduce is a paradigm which has two phases, the mapper phase, and the reducer phase. In the mapper, the input is given in the form of key-value pair. The output of the Mapper is fed to the reducer as input. The reducer runs only after the Mapper is over. The reducer too takes input in key-value format, and the output of reducer is the final output.
- * Steps in Map Reduce
 - The map takes data in the form of pairs and returns a list of $\langle \text{key}, \text{value} \rangle$ pairs. The keys will not be unique in this case.
 - Using the output of map, sort and shuffle are applied by the Hadoop architecture. This sort and shuffle acts on these list of $\langle \text{key}, \text{value} \rangle$ pairs and sends out unique keys and a list of values associated with this unique key $\langle \text{key}, \text{list}(\text{values}) \rangle$.
 - An output of sort and shuffle sent to the reducer phase. The reducer performs a defined function on a list of values for unique keys, and final output $\langle \text{key}, \text{value} \rangle$ will be stored / displayed.
- * Sort and shuffle
The sort and shuffle occur on the output of Mapper and before the reducer. When the Mapper task is complete, the results are sorted by key, partitioned if there are multiple reducers, and then written to disk. Using the input from each Mapper $\langle k_2, v_2 \rangle$, we collect all the values for each unique key k_2 . The output from the



shuffle phase in the form of $\langle k_2, \text{list}(V_2) \rangle$ is sent as input to reducer phase.

* Map reduce word count example

In Map reduce word count example, we find the frequency of each word. Here, the role of Mapper is to Map the keys to existing values and the role of Reducer is to aggregate the keys of common values. So, everything is represented in the form of key-value pair.



Conclusion :

Thus we have successfully imple the word count program in java