MapReduce is framework using which we can write applicatio
to process huge amount of data, in parallel, on large cluster of commodity hardware in a reliable manner.

What is MapReduce?

- MapReduce is processing technique and program model for distributed computing based on java.

- MapReduce paradigm is based on sending the computer to write where the data resides.

There are 2 stages in MapReduce

Stage 1 :- Map
Stage 2 :- Reduce
Both Map and Reduce only works on (key, value) pair

Map stage :- The map or mapper job is to process the input data.

- Generally the input data is in the form of file or directory and is stored in the HDFS.

- The input file is passed to the mapper function line by li

- The mapper process the data and creates several small chunks of data.

PAGE No.

DATE / /

Reduce Stage:- This stage is the combination of the shuffle stage and the Reduce stage.

- The Reducer's job is to process that data comes from the mapper.

- After processing, it produces a new set of output, which will be stored in the HDFS.

What is (key, value)?

Key	Value
Id	101
Name	Ram
Designation	Developer

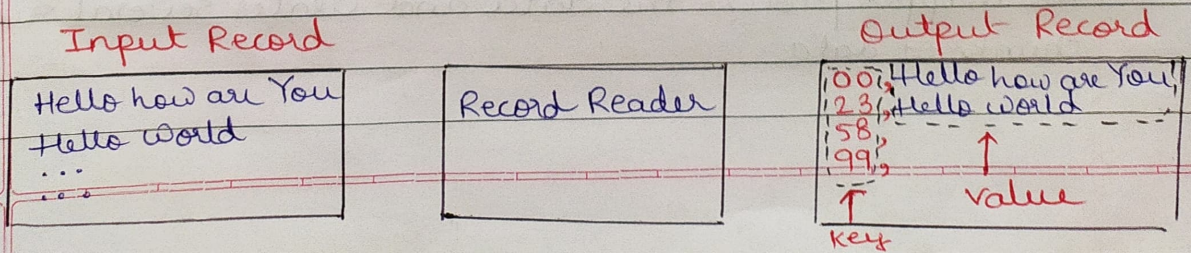
input　　　　　　　　output

<K,V> --→ M --→ <K,V> --→ R --→ <K,V>

　　　　　　　　　　input　　　　　　　　output

Record Reader

- The role of Record Reader is to convert each input line into (key, value) pair suitable for reading by Mapper.



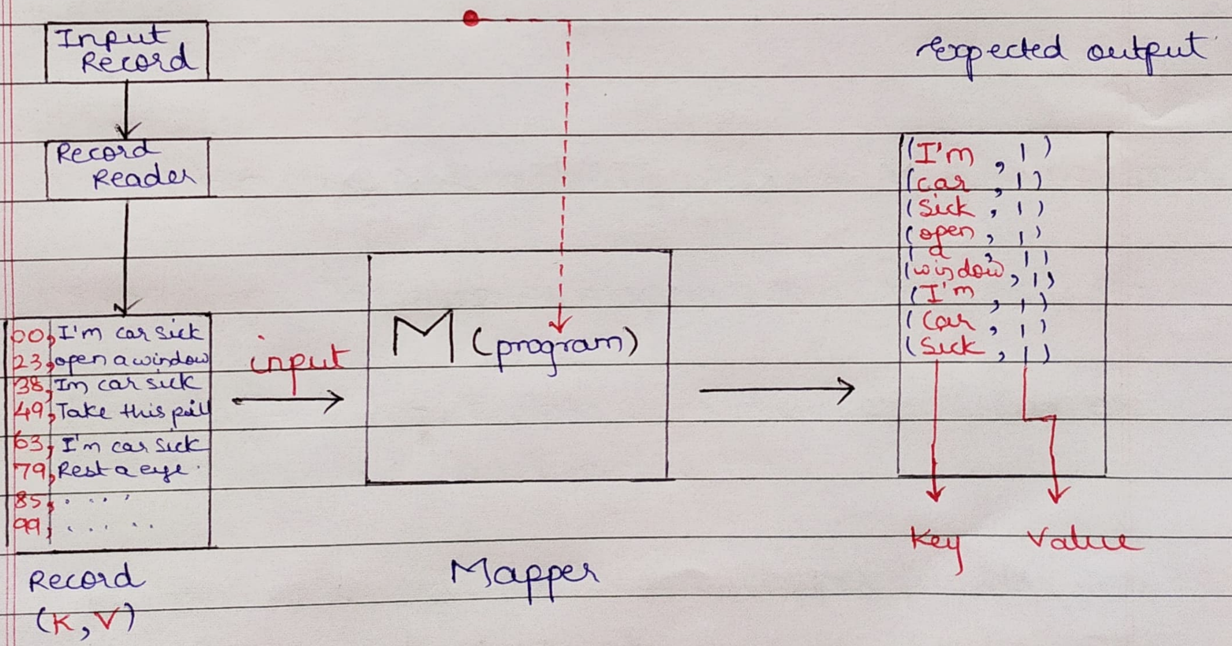
- During a MapReduce job, Hadoop sends the Map and Reduce task to the appropiate server in the cluster.

- The frameworks manages all the details of data-passing such as issuing task, verifying task completion, and copy data around the cluster between the nodes

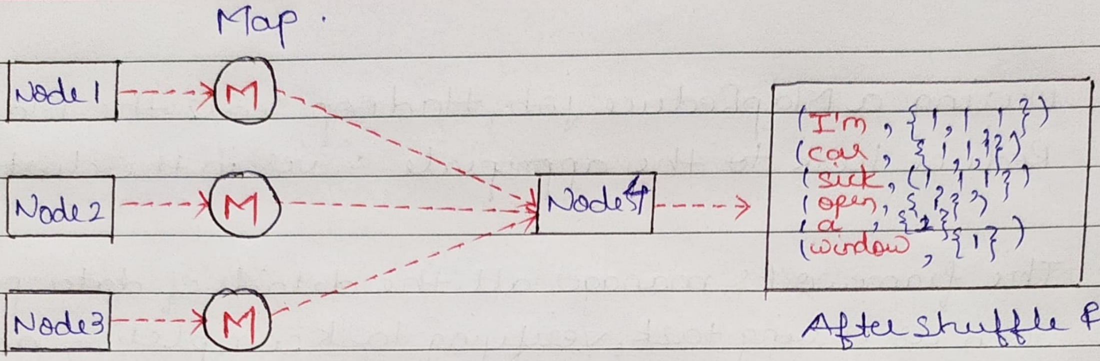
- Most of the computing takes places on nodes with the data on local disks that reduces the network traffic.

- After completion of the given tasks, the cluster collects and reduces the data to from an appropiate result, and sends it back to the Hadoop server.

How Mapper Reducers Works.



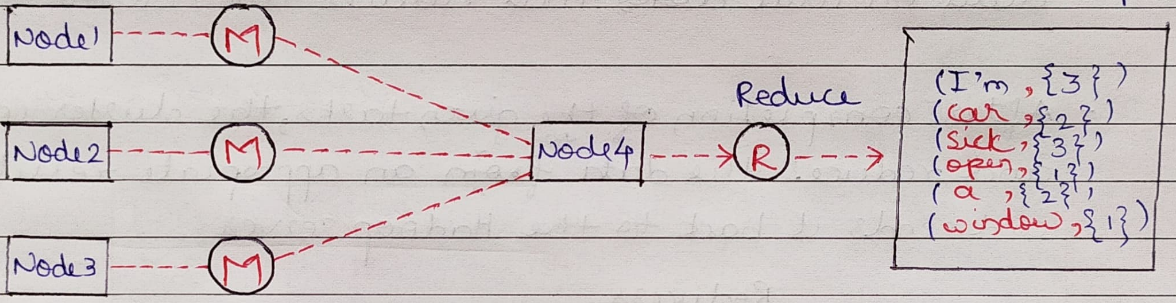
Map .



(I'm , {1, 1 , 1})
(car , {1, 1, 1})
(sick , (1, 1, 1})
(open, {1})
(a , {2})
(window , {1})

After shuffle & sort

Final output



(I'm , {3})
(car , {2})
(sick , {3})
(open , {1})
(a , {2})
(window , {1})