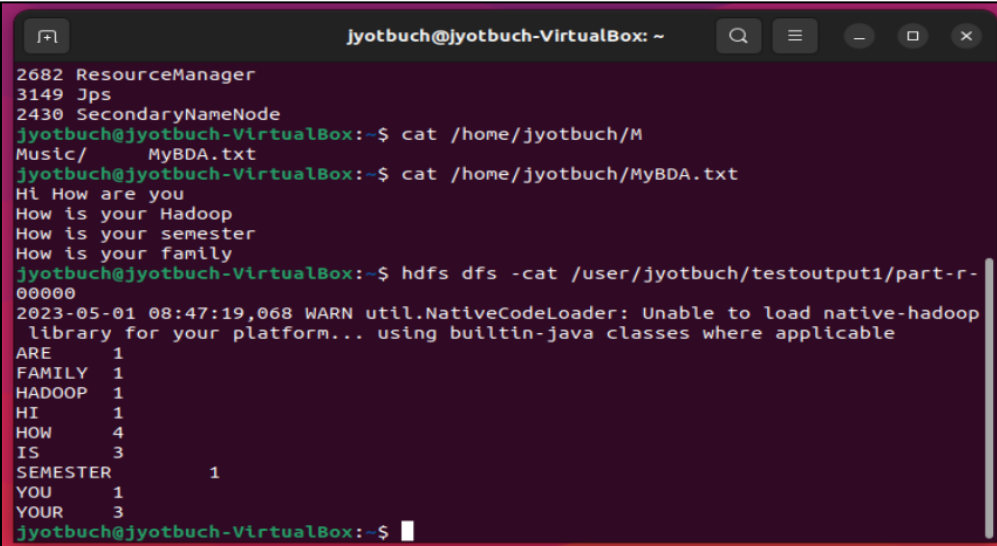


Assignment 2	
Name - Jyot Buch	Roll No. - 09
Date of Submission - 01/05/2023	Class - A (2023 Batch)

1. Discuss the functions of using Driver, Mapper and Reducer in Hadoop. You can take an example of word count problem. Support your answer with screenshots of your console.



```
jyotbuch@jyotbuch-VirtualBox: ~  
2682 ResourceManager  
3149 Jps  
2430 SecondaryNameNode  
jyotbuch@jyotbuch-VirtualBox:~$ cat /home/jyotbuch/M  
Music/      MyBDA.txt  
jyotbuch@jyotbuch-VirtualBox:~$ cat /home/jyotbuch/MyBDA.txt  
Hi How are you  
How is your Hadoop  
How is your semester  
How is your family  
jyotbuch@jyotbuch-VirtualBox:~$ hdfs dfs -cat /user/jyotbuch/testoutput1/part-r-  
00000  
2023-05-01 08:47:19,068 WARN util.NativeCodeLoader: Unable to load native-hadoop  
library for your platform... using builtin-java classes where applicable  
ARE      1  
FAMILY   1  
HADOOP   1  
HI        1  
HOW      4  
IS        3  
SEMESTER      1  
YOU       1  
YOUR      3  
jyotbuch@jyotbuch-VirtualBox:~$
```

The Driver, Mapper, and Reducer components in Hadoop work together to process and analyze large amounts of data.

Suppose we have a large text file, and we want to count the occurrences of each word in the file. Here's how we can use Hadoop to solve this problem:

1. Driver: The Driver is responsible for managing the overall execution of the Hadoop job. In this case, the Driver specifies the input file path, output file path, and the classes for the Mapper and Reducer. It also sets the number of Mapper and Reducer tasks.
2. Mapper: The Mapper is responsible for processing the input data and generating intermediate key-value pairs. In this case, the Mapper reads each line of the input file and tokenizes the words. For each word, the Mapper outputs a key-value pair, where the key is the word and the value is 1.
3. Reducer: The Reducer is responsible for processing the intermediate key-value pairs generated by the Mapper and producing the final output. In this case, the Reducer

	receives the key-value pairs generated by the Mapper and aggregates the values for each key. The output of the Reducer is a set of key-value pairs, where the key is the word and the value is the count of occurrences of the word.
2.	Explain YARN model and give its comparison with that of Hadoop
	<p>YARN is a framework for cluster resource management and job scheduling in Hadoop, introduced in Hadoop 2.0. It separates the responsibilities of resource management and job scheduling/monitoring into two separate daemons: the ResourceManager and the NodeManager. The ResourceManager is responsible for managing cluster resources and scheduling applications, while the NodeManager is responsible for launching and monitoring containers on each individual node. This separation allows for more scalability, multi-tenancy, and flexibility than the traditional Hadoop model.</p> <p>The traditional Hadoop model had a single point of failure and could become a bottleneck as the size of the cluster grew. With YARN, the ResourceManager is distributed and scalable, allowing for larger clusters and more concurrent applications. YARN also supports multi-tenancy, allowing multiple applications to run simultaneously on the same cluster, and allows for different processing engines to run on the same cluster, allowing for more diverse workloads and greater efficiency. Overall, the YARN model provides a more efficient and flexible framework for cluster resource management and job scheduling than the traditional Hadoop model.</p>