

LAB REPORT

Exercise-1

By

Ambuj Mishra

202116003

Big Data & Large-Scale Computing

DA-IICT, Gandhinagar

11-February-2022

1st Question:

The first question was about exploring the various syntax and methods that are followed while using Java. I have used textbook (not yet completed all chapters from 2 to 6) and resources from YouTube along with that. I previously have never worked on Java so below are a few learnings that I can summarize:

- i. I have got to know about the basic structure of a Java program that what is a class, what is a function, what is a Java application and how to create modules and classes inside a Java application.
- ii. I have also learned about variables, keywords, and control statements in Java.
- iii. I also went through different data structures in a Java code and type casting of any variable from one data structure to another.
- iv. I also studied about different operators, their precedence and associativity and went through the iterative statements and jump statements.

Shortcomings: Although the concept of OOPS is at the very start of chapter 2. But still, it is not clear to me completely. I am still working on that part. I am also working on fully understanding the concepts of access modifiers (i.e., Public, protected, private and default).

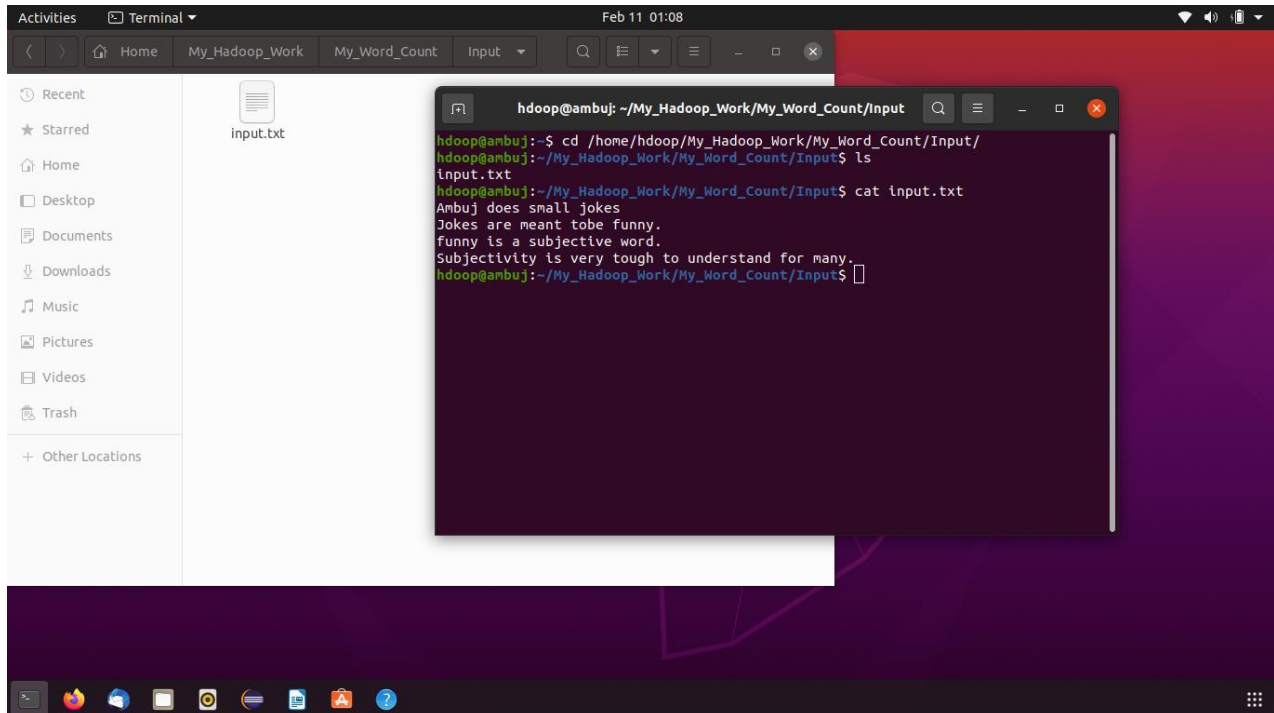
2nd Question:

The second question deals with the proper installation of Hadoop and integrating it with a Java code editor (I have used eclipse for this) and then running the word count code on it. Below are my observations and learnings for this task:

- i. I am using MacOS currently and tried installing Hadoop on it. It was working on a single node cluster properly but was not able to run resource manager on it. Hadoop distributed file system (hdfs) was also not working on it.
- ii. I installed ubuntu on a different system and even configured all settings of hdfs and yarn xml files.

- iii. I realized that providing proper paths of Java folder to the environment file is extremely critical as it will help in creating a mutual connection between Java and Hadoop.

Input File



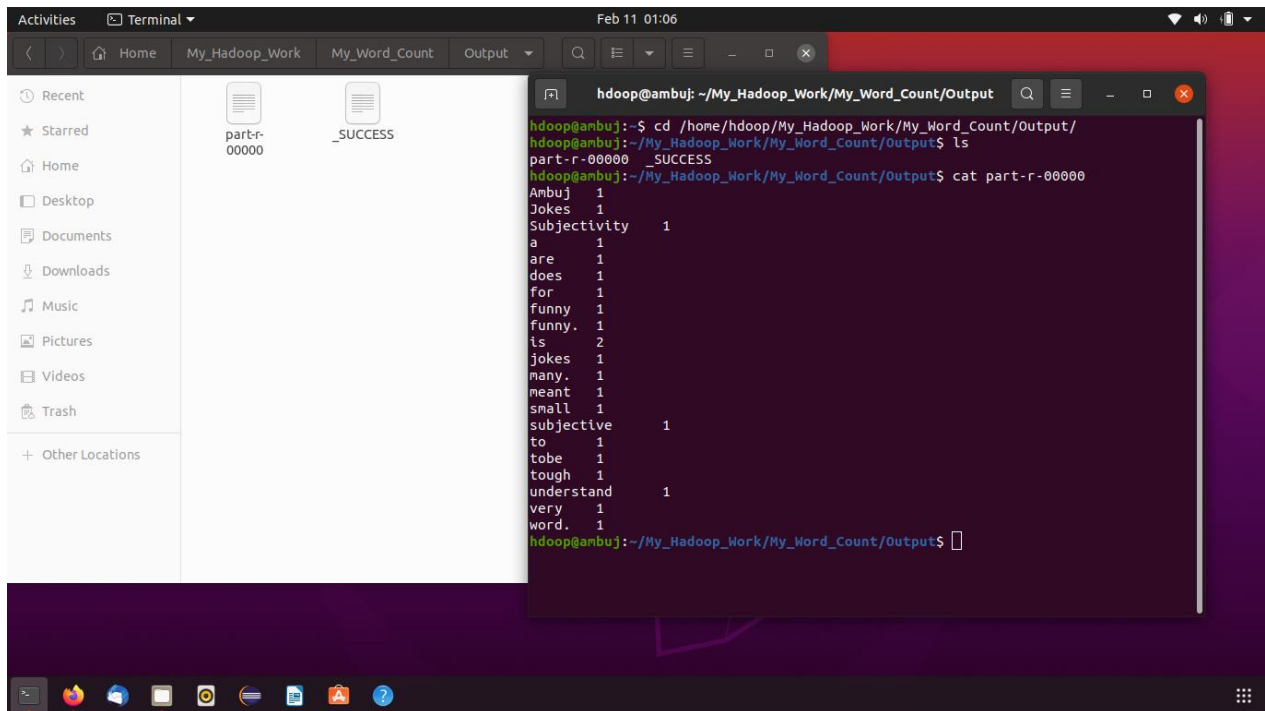
The screenshot shows a Linux desktop environment. On the left, a file manager window displays the contents of the `~/My_Hadoop_Work/My_Word_Count` directory, showing a file named `input.txt`. In the center, a terminal window is open, showing the following commands and output:

```
hadoop@ambuj: ~/My_Hadoop_Work/My_Word_Count/Input
hadoop@ambuj:~$ cd /home/hadoop/My_Hadoop_Work/My_Word_Count/Input/
hadoop@ambuj:~/My_Hadoop_Work/My_Word_Count/Input$ ls
input.txt
hadoop@ambuj:~/My_Hadoop_Work/My_Word_Count/Input$ cat input.txt
Ambuj does small jokes
Jokes are meant to be funny.
funny is a subjective word.
Subjectivity is very tough to understand for many.
hadoop@ambuj:~/My_Hadoop_Work/My_Word_Count/Input$
```

- iv. I tried using map-reduce example already present in the Hadoop directory, but it was not working. Then after a lot of research I found out that since I have configured all the hdfs and yarn settings, so I need to execute map-reduce algorithm using the hdfs system method.
- v. I tried to run word count code using hdfs and it started working and produced results as expected.
- vi. After the demonstration of Hadoop's integration with eclipse, I also tried using it, but it did not work initially because I was using .jar files only from map-reduce and common folders.

- vii. After including all the .jar files in the Java project, the map-reduce word count code started working on eclipse as well.

Output File



```
hadoop@ambuj: ~/My_Hadoop_Work/My_Word_Count/Output
hadoop@ambuj:~$ cd /home/hadoop/My_Hadoop_Work/My_Word_Count/Output/
hadoop@ambuj:~/My_Hadoop_Work/My_Word_Count/Output$ ls
part-r-00000  _SUCCESS
hadoop@ambuj:~/My_Hadoop_Work/My_Word_Count/Output$ cat part-r-00000
Anbuji 1
Jokes 1
Subjectivity 1
a 1
are 1
does 1
for 1
funny 1
funny. 1
is 2
jokes 1
many. 1
meant 1
small 1
subjective 1
to 1
tobe 1
tough 1
understand 1
very 1
word. 1
hadoop@ambuj:~/My_Hadoop_Work/My_Word_Count/Output$
```

Shortcomings: I am still very new to working with Hadoop. After the initial few lectures, I realized that following textbook is the best way to move through this course. I am currently doing the same and will try to continue that in future as well.
