Student Name	Aditya Wandhekar
SRN No	202201449
Roll No	34
Program	AIDS
Year	Third Year
Division	A
Subject	Big Data Analytics Lab
Assignment No	Five

5) Read and write text files to HDFS with Spark


```python
!pip install pyspark
from pyspark import SparkConf, SparkContext

# Initialize Spark configuration
conf = SparkConf().setAppName("MyApp").setMaster("local")

# Create SparkContext
sc = SparkContext(conf=conf)

# Now you can read the text file
lines = sc.textFile(r"\usr\bin\Documents\5th Sem\BDA\Cloudera\Hello.txt")

# Optionally, you can do something with the lines
print(lines.collect())
```

```python
lines.count()
```

```python
words=lines.flatMap(lambda line : line.split(""))
```

```python
tuples = words.map(lambda : (word,1))
```

```python
counts = tuplees.reduceByKey(lambda a,b(a+b))
```

```python
In [6]: counts.coalesce(1).saveAsTextFile('hdfs:/user/cloudera/wordcount/outputDir')
```