In one terminal,

diyabhat@LAPTOP-L0SC731M:~$ nc -lk 9999

In another terminal,

* diyabhat@LAPTOP-L0SC731M:~$ java -version

Command 'java' not found, but can be installed with:

sudo apt install openjdk-8-jre-headless # version 8u442-b06~us1-0ubuntu1~22.04

* diyabhat@LAPTOP-L0SC731M:~$ sudo apt install openjdk-8-jre-headless
* diyabhat@LAPTOP-L0SC731M:~$ java -version

openjdk version "1.8.0\_442"

* diyabhat@LAPTOP-L0SC731M:~$ cd /mnt/c/spark-3.5.5/bin
* diyabhat@LAPTOP-L0SC731M:/mnt/c/spark-3.5.5/bin$ ./spark-shell -i /mnt/c/scala/wordCount.scala

scala> import org.apache.spark.sql.SparkSession

scala> import org.apache.spark.sql.functions.\_

object WordCountStreaming {

def main(args: Array[String]): Unit =

{

// Initialize SparkSession

val spark = SparkSession.builder

.appName("StructuredNetworkWordCount")

.master("local[\*]")

.getOrCreate()

// Read streaming data from a socket

val lines = spark.readStream

.format("socket")

.option("host", "localhost")

.option("port", 9999)

.load()

import spark.implicits.\_

// Split lines into words

val words = lines.as[String].flatMap(\_.split(" "))

// Group words and count occurrences

val wordCounts = words.groupBy("value").count()

// Write the output to the console

val query = wordCounts.writeStream

.outputMode("complete")

.format("console")

.start()

// Await termination of the streaming query

query.awaitTermination()

}

}

scala> WordCountStreaming.main(Array())

diyabhat@LAPTOP-L0SC731M:~$ stop-all.sh

WARNING: Stopping all Apache Hadoop daemons as diyabhat in 10 seconds.

WARNING: Use CTRL-C to abort.

Stopping namenodes on [localhost]

Stopping datanodes

Stopping secondary namenodes [LAPTOP-L0SC731M]

Stopping nodemanagers

Stopping resourcemanager

start-all.sh

WARNING: Attempting to start all Apache Hadoop daemons as diyabhat in 10 seconds.

WARNING: This is not a recommended production deployment configuration.

WARNING: Use CTRL-C to abort.

Starting namenodes on [localhost]

Starting datanodes

Starting secondary namenodes [LAPTOP-L0SC731M]

Starting resourcemanager

Starting nodemanagers

diyabhat@LAPTOP-L0SC731M:~$ jps

2131 ResourceManager

2621 Jps

1901 SecondaryNameNode

diyabhat@LAPTOP-L0SC731M:~$ start-dfs.sh

c

diyabhat@LAPTOP-L0SC731M:~$ find /home/diyabhat -type f -name "hadoop-diyabhat-namenode-\*.log"

/home/diyabhat/hadoop-3.4.1/logs/hadoop-diyabhat-namenode-LAPTOP-L0SC731M.log

diyabhat@LAPTOP-L0SC731M:~$ kill -9 12925

diyabhat@LAPTOP-L0SC731M:~$ hadoop fs -put input.txt /user7/input

put: `input.txt': No such file or directory

diyabhat@LAPTOP-L0SC731M:~$ echo "This is a sample input text for Hadoop." > input.txt

diyabhat@LAPTOP-L0SC731M:~$ hadoop fs -put input.txt /user7/input

diyabhat@LAPTOP-L0SC731M:~$ hadoop jar /mnt/d/College/Sem6/DSBDAL/Hadoop/example1.jar hadoopPrac.wordcount /user7/input /user7/output

diyabhat@LAPTOP-L0SC731M:~$ nano input.txt

diyabhat@LAPTOP-L0SC731M:~$ cat input.txt

This is a sample input text for Hadoop input file it is Hadoop.

Diya

Kalyani

Diya

Sakshi

Vidhi

Koul

Kalyani

Sanjay

Patil

Sakshi

diyabhat@LAPTOP-L0SC731M:~$ hadoop fs -mkdir -p /user7/input

diyabhat@LAPTOP-L0SC731M:~$ hadoop fs -put input.txt /user7/input

diyabhat@LAPTOP-L0SC731M:~$ hadoop fs -ls /user7/input

Found 1 items

-rw-r--r-- 3 diyabhat supergroup 128 2025-04-26 12:27 /user7/input/input.txt

diyabhat@LAPTOP-L0SC731M:~$

>

C:\Users\diyab\eclipse-workspace\WordCount\src\hadoopPrac>javac -cp "C:\Users\diyab\Desktop\hadoop-jars\\*" -d C:\Users\diyab\eclipse-workspace\WordCount\bin wordcount.java

import org.apache.spark.sql.SparkSession

import org.apache.spark.sql.functions.\_

object WordCountStreaming{

def main(args:Array[String]):Unit={

val spark=SparkSession

.builder

` .appName(“StructuredNetworkWordCount”)

.master(“local[\*]”)

.getOrCreate()

val lines=spark

.readStream

.format(socket)

.option(“host”,”localhost”)

.option(“port”,9999)

.load()

import spark.implicits.\_

val words=lines

.as[String]

.flatMap(\_.split(“ “))

val wordCounts=words.groupBy(“value”).count()

val query=wordCounts

.writeStream

.outputMode(“complete”)

.format(“console”)

.start()

query.await.Termination()

}

}