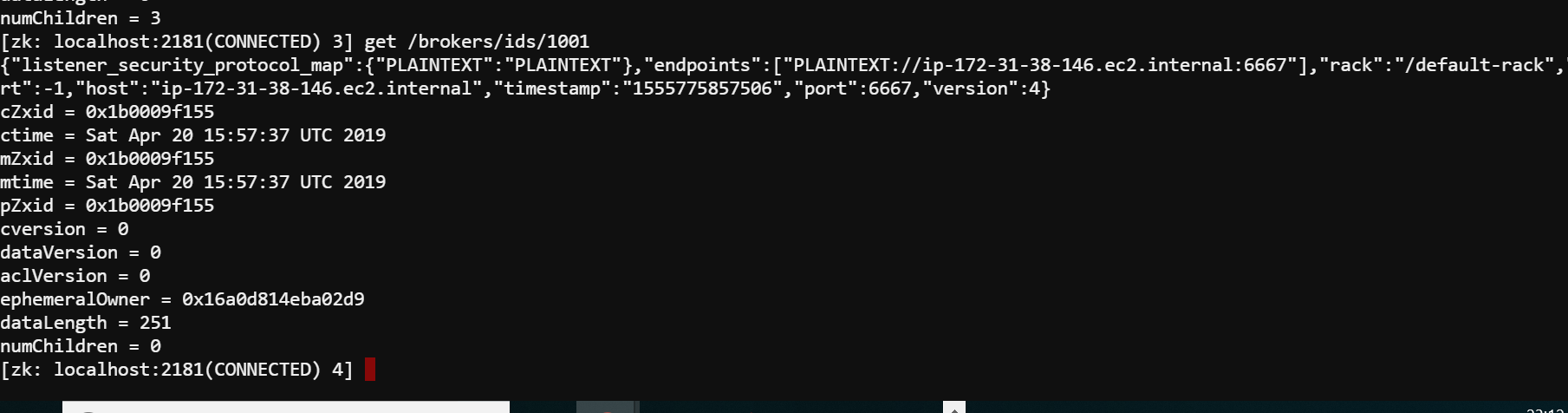
**Purpose: creating kafka topics , push messages and check the count of unique words every 2 seconds**

**Step1**: Login to zookeeper console and find out the broker local host IP addresses , since Kafka brokers communicate through zookeeper.



The Ip address of broker 1001 is **ip-172-31-38-146.ec2.internal**

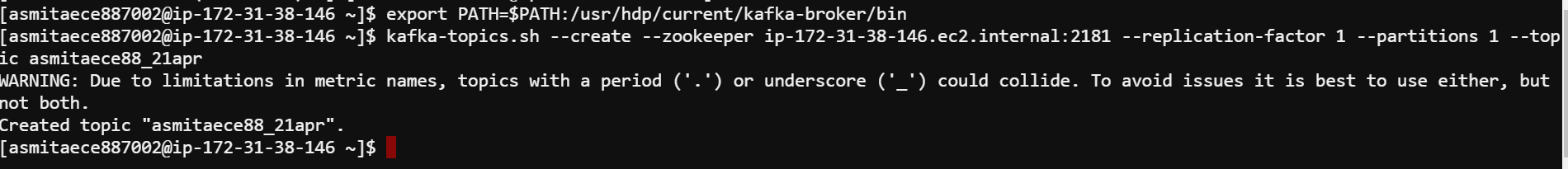
**Step2: creation of kafka topic**

# Include Kafka binaries in the path. HDP includes the kafka and installs at /usr/hdp/current/kafka-broker

export PATH=$PATH:/usr/hdp/current/kafka-broker/bin

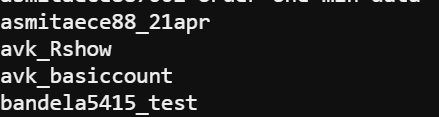
# Create the topic named as asmitaece88\_21apr

kafka-topics.sh --create --zookeeper ip-172-31-38-146.ec2.internal:2181 --replication-factor 1 --partitions 1 --topic asmitaece88\_21apr



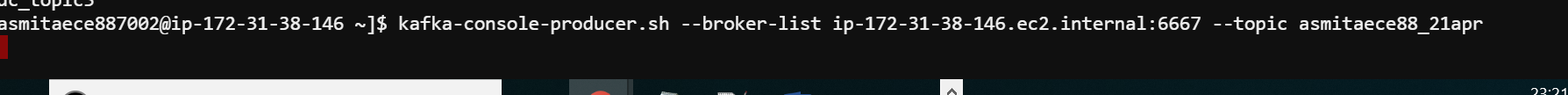
#Check if topic created

kafka-topics.sh --list --zookeeper ip-172-31-38-146.ec2.internal:2181

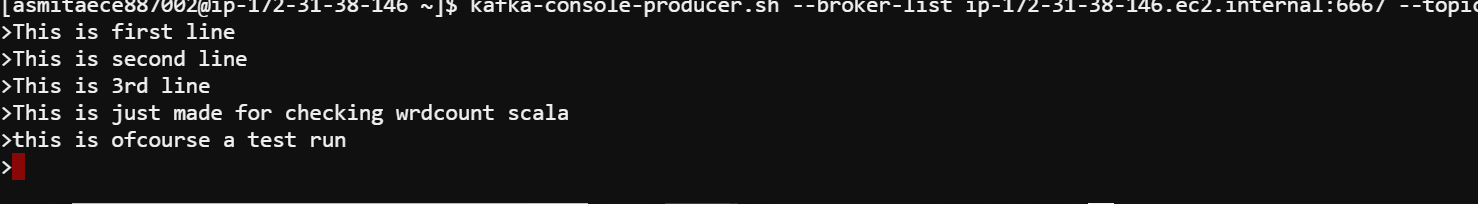


**Step3 : Interacting with kafka – produce**

kafka-console-producer.sh --broker-list ip-172-31-38-146.ec2.internal:6667 --topic asmitaece88\_21apr



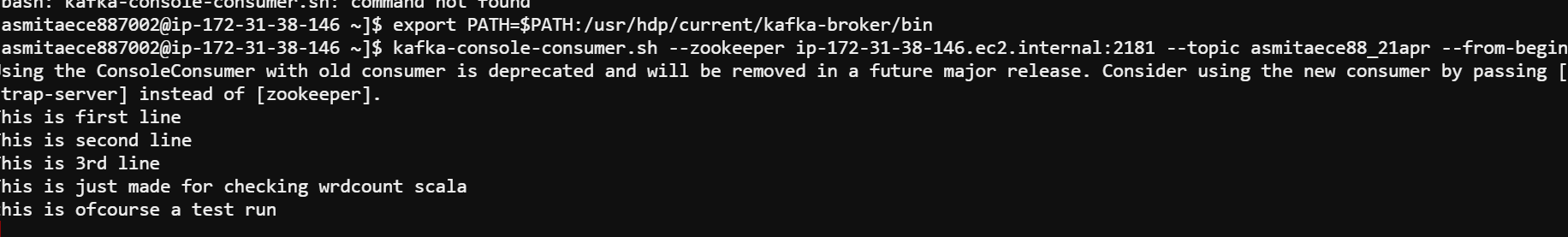
Next push messages as follows :



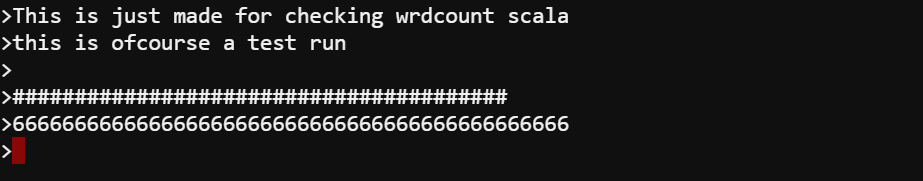
Check the kafka consumer :

export PATH=$PATH:/usr/hdp/current/kafka-broker/bin

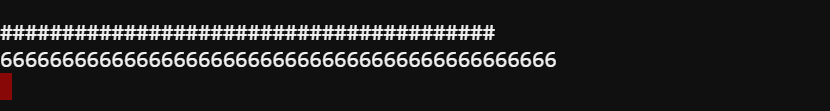
kafka-console-consumer.sh --zookeeper ip-172-31-38-146.ec2.internal:2181 --topic asmitaece88\_21apr --from-beginning



Again , typing some new messages in kafka producer console :



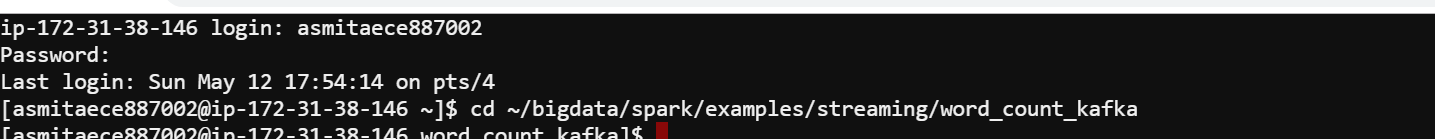
Now checking the Kafka consumer console , we get the messages automatically populated in consumer



**Step3 : Spark streaming**

#Leave the producer prompt open and open a new terminal

# If you are already not in the project folder, please change directory



# Build the package

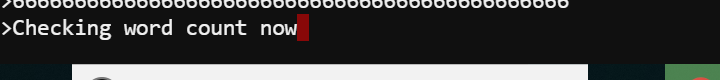
# Note that there are only two files in the project: build.sbt and one scala program

# Using scala build tool, first clean the build and then package it

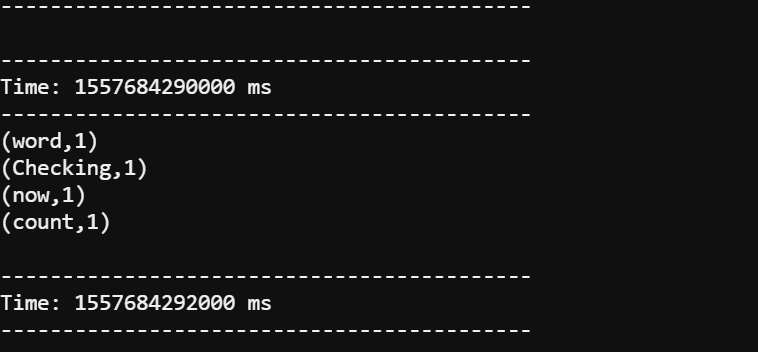
# Notice that "&&" is the usual unix operator to chain commands

sbt clean && sbt package

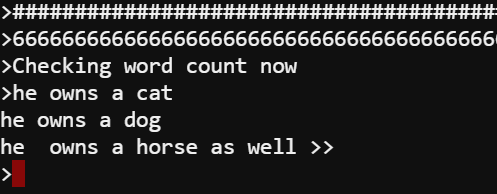
Next Push the messages in producer console and press enter



In the spark streaming web console , the word counts of the new messages are obtained every 2 seconds



In the producer console , entre new messages :



In the spark streaming console :

