```
onsole-consumer.sh --topic lesson2_topic_student910_14 --from-beginning --bootstrap-server bigda
taanalytics2-worker-shdpt-v31-1-4:6667
4 message
```

[student910\_14@bigdataanalytics2-worker-shdpt-v31-1-0 ~]\$ /usr/hdp/3.1.4.0-315/kafka/bin/kafka-c

```
[student910_14@bigdataanalytics2-worker-shdpt-v31-1-0 ~]$ /usr/hdp/3.1.4.0-315/kafka/bin/kafka-c
onsole-producer.sh --topic lesson2_topic_student910_14 --broker-list bigdataanalytics2-worker-sh
dpt-v31-1-5:6667
>4 message
```

[student910\_14@bigdataanalytics2-worker-shdpt-v31-1-0 ~]\$ /usr/hdp/3.1.4.0-315/kafka/bin/kafka-r

<u>un-class.sh kafka.tools.</u>GetOffsetShell --broker-list bigdataanalytics2-worker-shdpt-v31-1-5:6667

--topic lesson2\_topic\_student910\_14

lesson2\_topic\_student910\_14:0:5

[student910\_14@bigdataanalytics2-worker-shdpt-v31-1-0 ~]\$ /usr/hdp/3.1.4.0-315/kafka/bin/kafka-r

un-class.sh kafka.tools.GetOffsetShell --broker-list bigdataanalytics2-worker-shdpt-v31-1-5:6667

--topic lesson2\_topic\_student910\_14 --offsets 1 | awk -F ":" '{sum += \$3} END {print sum}'

```
version 2.4.7
Using Python version 2.7.5 (default, Nov 16 2020 22:23:17)
SparkSession available as 'spark'.
>>> checkpoint location = "/user/student910 14/tmp/spark-stream-chkpnt/"
>>> spark = SparkSession \
   .builder \
        .appName("FileSourceApp") \
        .master("local[2]") \
        .getOrCreate()
>>> spark.sparkContext.setLogLevel("ERROR")
>>> file df = spark \
        .readStream \
        .format("csv") \
        .schema("kev INT, value STRING") \
        .options(maxFilesPerTrigger=1) \
        .load("/user/student910_14/tmp/*")
>>> file_stream = file_df \
        .writeStream \
        .format("console") \
        .trigger(processingTime="10 Seconds") \
        .option("checkpointLocation", checkpoint location) \
        .start()
>>> --
Batch: 0
| key|value|
+----+
|null| null|
Batch: 1
| key| value|
Inull| null|
    2 message
```

Welcome to

```
[student910_14@bigdataanalytics2-worker-shdpt-v31-1-0 ~]$ mkdir hw2
[student910_14@bigdataanalytics2-worker-shdpt-v31-1-0 ~]$ cd hw2
[student910_14@bigdataanalytics2-worker-shdpt-v31-1-0 hw2]$ echo "1 message" >> ./1.csv
[student910_14@bigdataanalytics2-worker-shdpt-v31-1-0 hw2]$ echo "2 message" >> ./2.csv
[student910_14@bigdataanalytics2-worker-shdpt-v31-1-0 hw2]$ hdfs dfs -put 1.csv /user/student910
_14/tmp
[student910_14@bigdataanalytics2-worker-shdpt-v31-1-0 hw2]$ echo "2,message" >> ./2.csv
[student910_14@bigdataanalytics2-worker-shdpt-v31-1-0 hw2]$ hdfs dfs -put 2.csv /user/student910
_14/tmp
```