Code Logic - Retail Data Analysis

Setting up with some Dependencies by importing dependencies with libraries.

```
import os
import sys
from pyspark.sql import SparkSession
from pyspark.sql.functions import *
from pyspark.sql.types import *
from pyspark.sql.functions import from_json
from pyspark.sql.window import Window
```

Initialising spark session and setting log level to error.

```
spark = SparkSession  \
        .builder  \
        .appName("spark-streaming")  \
        .getOrCreate()
spark.sparkContext.setLogLevel('ERROR')
```

Reading input data from Kafka mentioning the details of the Kafka broker, such as bootstrap server, port and topic name.

```
raw_order = spark  \
        .readStream  \
        .format("kafka")  \
        .option("kafka.bootstrap.servers","18.211.252.152:9092")  \
        .option("subscribe","real-time-project")  \
        .option("startingOffsets", "latest")  \
        .load()
```

Defining JSON schema of each order, using appropriate datatypes and StrucField in the case of the item attributes.

```
JSON_Schema = StructType() \
        .add("invoice_no", LongType()) \
         .add("country",StringType()) \
        .add("timestamp", TimestampType()) \
        .add("type", StringType()) \
        .add("total_items",IntegerType())\
        .add("is_order",IntegerType()) \
        .add("is_return",IntegerType()) \
        .add("items", ArrayType(StructType([
        StructField("SKU", StringType()),
        StructField("title", StringType()),
        StructField("unit_price", FloatType()),
        StructField("quantity", IntegerType())
        ])))
```

Reading the raw JSON data from Kafka as ‘order stream’ by casting it to string and storing it into the alias ‘data’.

```
order_stream =
raw_order.select(from_json(col("value").cast("string"),JSON_Schema).alias("data")).select("data.*")
```

Total Cost UDF - To calculate the total income from every invoice I needed to calculate the income from sale of each product, so I multiplied the unit price of the product with the quantity of the product purchased.

```
def total_cost(items,type):
   total_price = 0
   for item in items:
       total_price = total_price + item['unit_price'] * item['quantity']
   if type=="RETURN":
       return total_price * (-1)
   else:
       return total_price
```

Total Items UDF - To calculate the number of products in every invoice I added the quantity ordered of each product in that invoice.

```
def total_item_count(items):
   total_count = 0
   for item in items:
       total_count = total_count + item['quantity']
   return total_count
```

Is Order UDF - To determine if invoice is for an order or not I used an if-else statement.

```
def is_a_order(type):
  if type=="ORDER":
      return 1
  else:
      return 0
```

Is Return UDF - To determine if invoice is for a return or not I used an if-else statement.

```
def is_a_return(type):
  if type=="RETURN":
      return 1
  else:
      return 0
```

Defining the UDFs by Converting the Python functions I defined earlier, and assigning the appropriate return datatype.

```
is_order = udf(is_a_order, IntegerType())
is_return = udf(is_a_return, IntegerType())
add_total_item_count = udf(total_item_count, IntegerType())
add_total_cost = udf(total_cost, FloatType())
```

Calculating the additional columns according to the required input values.

```
order_extended_stream = order_stream \
    .withColumn("total_items", add_total_item_count(order_stream.items)) \
    .withColumn("total_cost", add_total_cost(order_stream.items,order_stream.type)) \
    .withColumn("is_order", is_order(order_stream.type)) \
    .withColumn("is_return", is_return(order_stream.type))
```

Writing the summarised input values to console, using ‘append’ output method and applying truncate as false and setting the processing time to 1 minute.

```
order_query_console = order_extended_stream \
    .select("invoice_no", "country", "timestamp","type","total_items","total_cost","is_order","is_return") \
    .writeStream \
    .outputMode("append") \
    .format("console") \
    .option("truncate", "false") \
    .trigger(processingTime="1 minute") \
    .start()
```

Calculating time-based KPIs (Total sale volume, OPM, Rate of return, Average transaction size) having tumbling window of one minute and watermark of one minute.

```
agg_time = order_extended_stream \
    .withWatermark("timestamp","1 minutes") \
    .groupby(window("timestamp", "1 minute")) \
    .agg(sum("total_cost").alias("total_volume_of_sales"),
        avg("total_cost").alias("average_transaction_size"),
        avg("is_Return").alias("rate_of_return")) \
    .select("window.start","window.end","total_volume_of_sales","average_transaction_size","rate_of_return")
```

Writing the time-based KPIs data to HDFS - HDFS into JSON files for each one-minute window, using ‘append’ output mode, setting truncate as false, and specifying the HDFS output path for both the KPI files and for their checkpoints. Ten 1-minute window batches were taken.

```
agg_time_country = order_extended_stream \
    .withWatermark("timestamp", "1 minutes") \
    .groupBy(window("timestamp", "1 minutes"), "country") \
    .agg(sum("total_cost").alias("total_volume_of_sales"),
        count("invoice_no").alias("OPM"),
        avg("is_Return").alias("rate_of_return")) \
    .select("window.start","window.end","country", "OPM","total_volume_of_sales","rate_of_return")
```

Writing the the time-and-country-based KPIs data to HDFS into JSON files for each one-minute window, using ‘append’ output mode, setting truncate as false, and specifying the HDFS output path for both the KPI files and for their checkpoints. Ten 1-minute window batches were taken.

```
ByTime_country = agg_time_country.writeStream \
    .format("json") \
    .outputMode("append") \
    .option("truncate", "false") \
    .option("path", "time_countryKPI/") \
    .option("checkpointLocation", "time_countryKPI/cp/") \
    .trigger(processingTime="1 minutes") \
    .start()
```

Indicating Spark to await termination.

```
extendedOrderQuery.awaitTermination()
agg_time_country.awaitTermination()
agg_time.awaitTermination()
```

Console – output

I have created cluster and logging with hadoop with ppk file.

I have used ls command to check how many file in hadoop.

vi spark-streaming.py to open file in console to change.

Kafka Version using the following command.

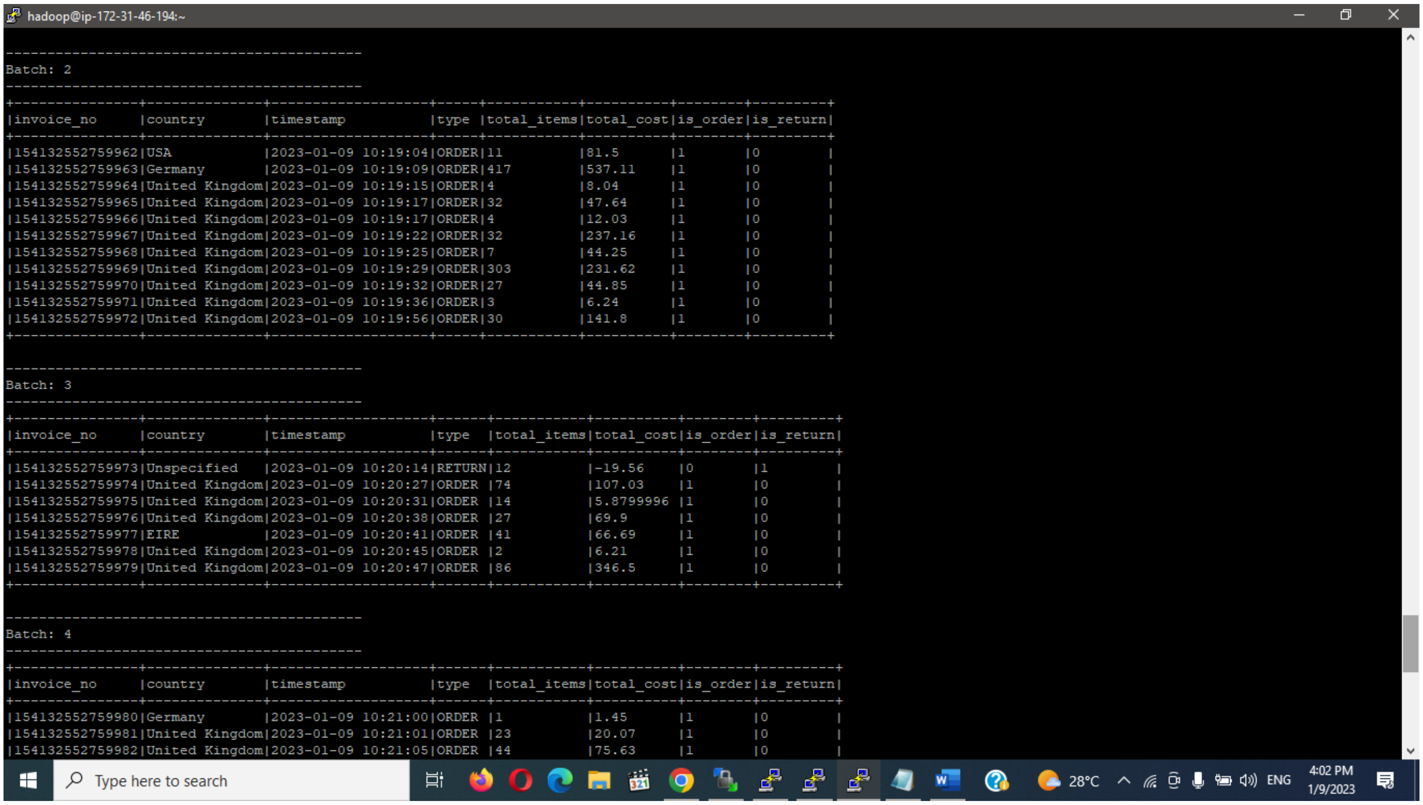
```
export SPARK_KAFKA_VERSION=0.10
```

I ran the spark2-submit command, specifying the jar and python file .

```
spark2-submit --jars spark-sql-kafka-0-10_2.11-2.3.0.jar spark-streaming.py 18.211.252.152:9092 real-time-project > console-output.txt
cat console-output.txt
```

To print output in one file

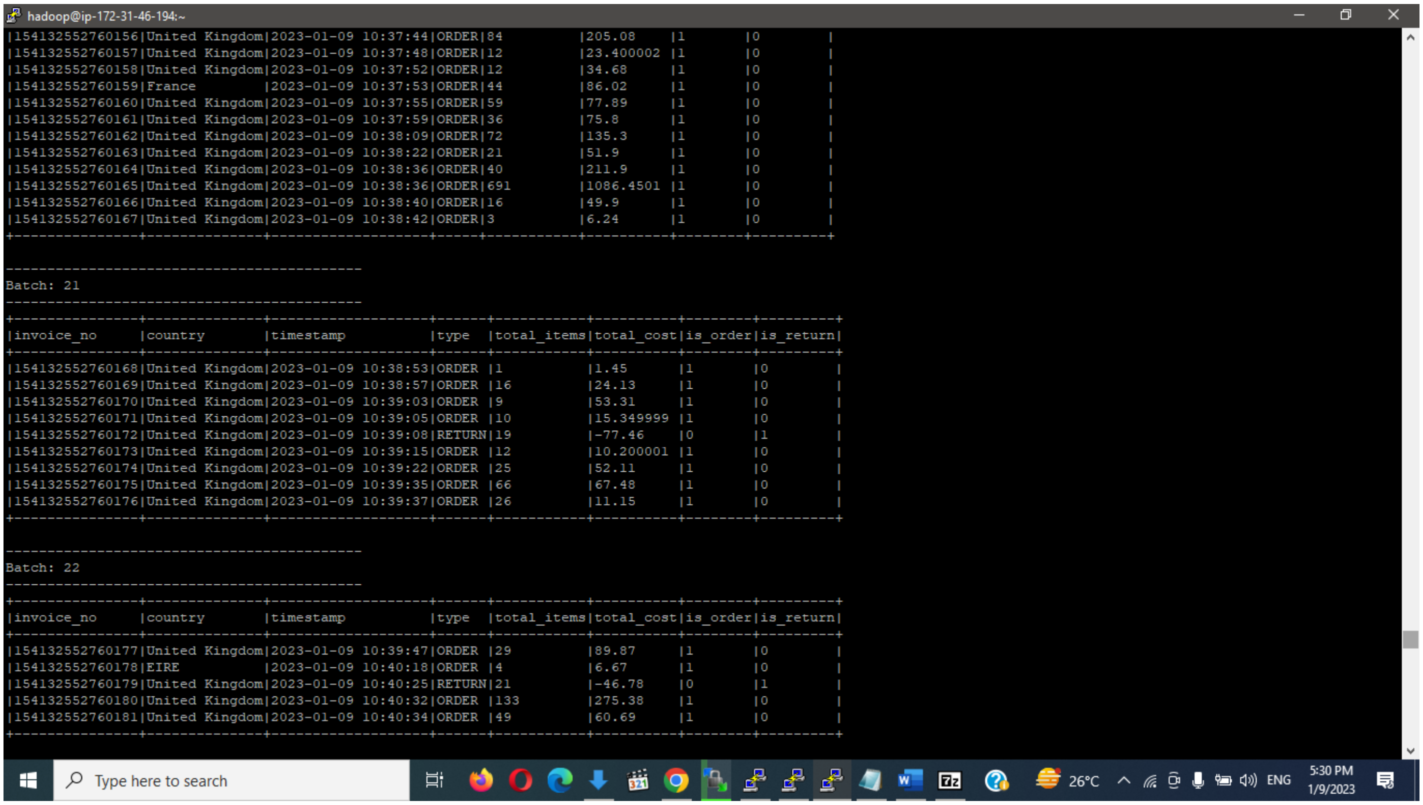
Example table - Final Summarised Input Values.



```
-------------------------------------------
Batch: 2
-------------------------------------------
+---------------+--------------+-------------------+-----+-----------+----------+--------+---------+
|invoice_no     |country       |timestamp          |type |total_items|total_cost|is_order|is_return|
+---------------+--------------+-------------------+-----+-----------+----------+--------+---------+
|154132552759962|USA           |2023-01-09 10:19:04|ORDER|11         |81.5      |1       |0        |
|154132552759963|Germany       |2023-01-09 10:19:09|ORDER|417        |537.11    |1       |0        |
|154132552759964|United Kingdom|2023-01-09 10:19:15|ORDER|4          |8.04      |1       |0        |
|154132552759965|United Kingdom|2023-01-09 10:19:17|ORDER|32         |47.64     |1       |0        |
|154132552759966|United Kingdom|2023-01-09 10:19:17|ORDER|4          |12.03     |1       |0        |
|154132552759967|United Kingdom|2023-01-09 10:19:22|ORDER|32         |237.16    |1       |0        |
|154132552759968|United Kingdom|2023-01-09 10:19:25|ORDER|7          |44.25     |1       |0        |
|154132552759969|United Kingdom|2023-01-09 10:19:29|ORDER|303        |231.62    |1       |0        |
|154132552759970|United Kingdom|2023-01-09 10:19:32|ORDER|27         |44.85     |1       |0        |
|154132552759971|United Kingdom|2023-01-09 10:19:36|ORDER|3          |6.24      |1       |0        |
|154132552759972|United Kingdom|2023-01-09 10:19:56|ORDER|30         |141.8     |1       |0        |
+---------------+--------------+-------------------+-----+-----------+----------+--------+---------+

-------------------------------------------
Batch: 3
-------------------------------------------
+---------------+--------------+-------------------+------+-----------+----------+--------+---------+
|invoice_no     |country       |timestamp          |type  |total_items|total_cost|is_order|is_return|
+---------------+--------------+-------------------+------+-----------+----------+--------+---------+
|154132552759973|Unspecified   |2023-01-09 10:20:14|RETURN|12         |-19.56    |0       |1        |
|154132552759974|United Kingdom|2023-01-09 10:20:27|ORDER |74         |107.03    |1       |0        |
|154132552759975|United Kingdom|2023-01-09 10:20:31|ORDER |14         |5.8799996 |1       |0        |
|154132552759976|United Kingdom|2023-01-09 10:20:38|ORDER |27         |69.9      |1       |0        |
|154132552759977|EIRE          |2023-01-09 10:20:41|ORDER |41         |66.69     |1       |0        |
|154132552759978|United Kingdom|2023-01-09 10:20:45|ORDER |2          |6.21      |1       |0        |
|154132552759979|United Kingdom|2023-01-09 10:20:47|ORDER |86         |346.5     |1       |0        |
+---------------+--------------+-------------------+------+-----------+----------+--------+---------+

-------------------------------------------
Batch: 4
-------------------------------------------
+---------------+--------------+-------------------+------+-----------+----------+--------+---------+
|invoice_no     |country       |timestamp          |type  |total_items|total_cost|is_order|is_return|
+---------------+--------------+-------------------+------+-----------+----------+--------+---------+
|154132552759980|Germany       |2023-01-09 10:21:00|ORDER |1          |1.45      |1       |0        |
|154132552759981|United Kingdom|2023-01-09 10:21:01|ORDER |23         |20.07     |1       |0        |
|154132552759982|United Kingdom|2023-01-09 10:21:05|ORDER |44         |75.63     |1       |0        |
```



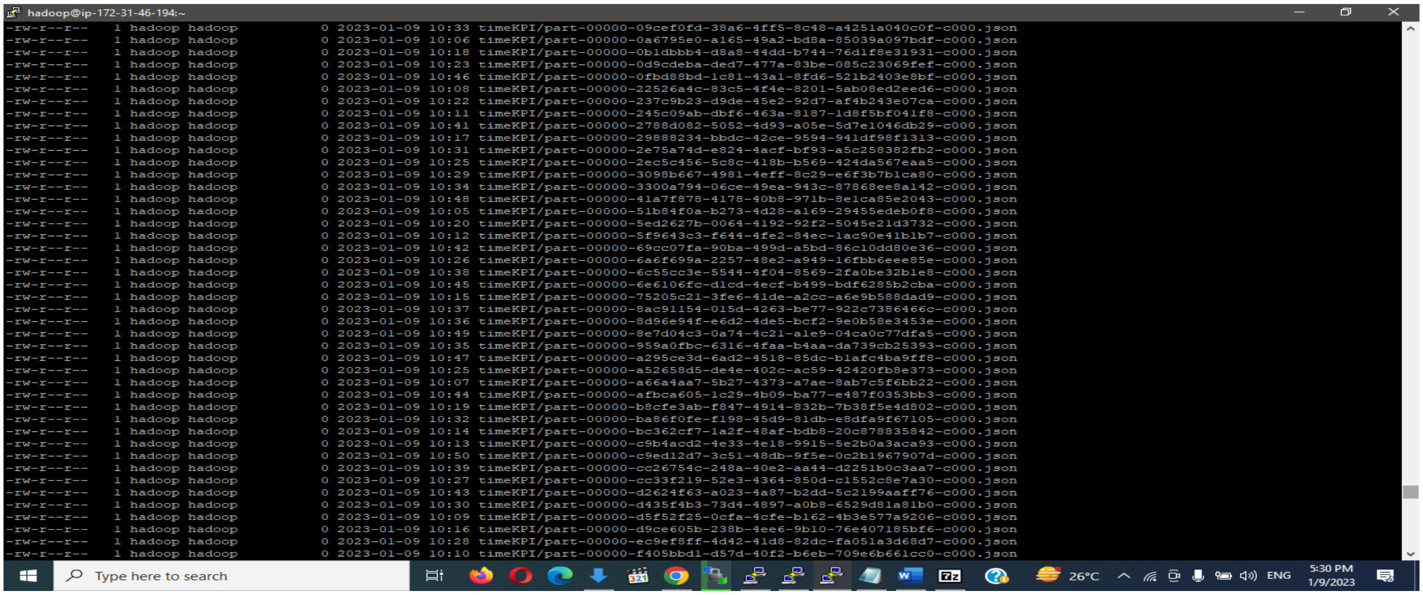
```
|15413255276015б|United Kingdom|2023-01-09 10:37:44|ORDER|84         |205.08    |1       |0        |
|154132552760157|United Kingdom|2023-01-09 10:37:48|ORDER|12         |23.400002 |1       |0        |
|154132552760158|United Kingdom|2023-01-09 10:37:52|ORDER|12         |34.68     |1       |0        |
|154132552760159|France        |2023-01-09 10:37:53|ORDER|44         |86.02     |1       |0        |
|154132552760160|United Kingdom|2023-01-09 10:37:55|ORDER|59         |77.89     |1       |0        |
|154132552760161|United Kingdom|2023-01-09 10:37:59|ORDER|36         |75.8      |1       |0        |
|154132552760162|United Kingdom|2023-01-09 10:38:09|ORDER|72         |135.3     |1       |0        |
|154132552760163|United Kingdom|2023-01-09 10:38:22|ORDER|21         |51.9      |1       |0        |
|154132552760164|United Kingdom|2023-01-09 10:38:36|ORDER|40         |211.9     |1       |0        |
|154132552760165|United Kingdom|2023-01-09 10:38:36|ORDER|691        |1086.4501 |1       |0        |
|154132552760166|United Kingdom|2023-01-09 10:38:40|ORDER|16         |49.9      |1       |0        |
|154132552760167|United Kingdom|2023-01-09 10:38:42|ORDER|3          |6.24      |1       |0        |
+---------------+--------------+-------------------+-----+-----------+----------+--------+---------+

-------------------------------------------
Batch: 21
-------------------------------------------
+---------------+--------------+-------------------+------+-----------+----------+--------+---------+
|invoice_no     |country       |timestamp          |type  |total_items|total_cost|is_order|is_return|
+---------------+--------------+-------------------+------+-----------+----------+--------+---------+
|154132552760168|United Kingdom|2023-01-09 10:38:53|ORDER |1          |1.45      |1       |0        |
|154132552760169|United Kingdom|2023-01-09 10:38:57|ORDER |16         |24.13     |1       |0        |
|154132552760170|United Kingdom|2023-01-09 10:39:03|ORDER |9          |53.31     |1       |0        |
|154132552760171|United Kingdom|2023-01-09 10:39:05|ORDER |10         |15.349999 |1       |0        |
|154132552760172|United Kingdom|2023-01-09 10:39:08|RETURN|19         |-77.46    |0       |1        |
|154132552760173|United Kingdom|2023-01-09 10:39:15|ORDER |12         |10.200001 |1       |0        |
|154132552760174|United Kingdom|2023-01-09 10:39:22|ORDER |25         |52.11     |1       |0        |
|154132552760175|United Kingdom|2023-01-09 10:39:35|ORDER |66         |67.48     |1       |0        |
|154132552760176|United Kingdom|2023-01-09 10:39:37|ORDER |26         |11.15     |1       |0        |
+---------------+--------------+-------------------+------+-----------+----------+--------+---------+

-------------------------------------------
Batch: 22
-------------------------------------------
+---------------+--------------+-------------------+------+-----------+----------+--------+---------+
|invoice_no     |country       |timestamp          |type  |total_items|total_cost|is_order|is_return|
+---------------+--------------+-------------------+------+-----------+----------+--------+---------+
|154132552760177|United Kingdom|2023-01-09 10:39:47|ORDER |29         |89.87     |1       |0        |
|154132552760178|EIRE          |2023-01-09 10:40:18|ORDER |4          |6.67      |1       |0        |
|154132552760179|United Kingdom|2023-01-09 10:40:25|RETURN|21         |-46.78    |0       |1        |
|154132552760180|United Kingdom|2023-01-09 10:40:32|ORDER |133        |275.38    |1       |0        |
|154132552760181|United Kingdom|2023-01-09 10:40:34|ORDER |49         |60.69     |1       |0        |
+---------------+--------------+-------------------+------+-----------+----------+--------+---------+
```

I checked HDFS to make sure the KPI files were present.

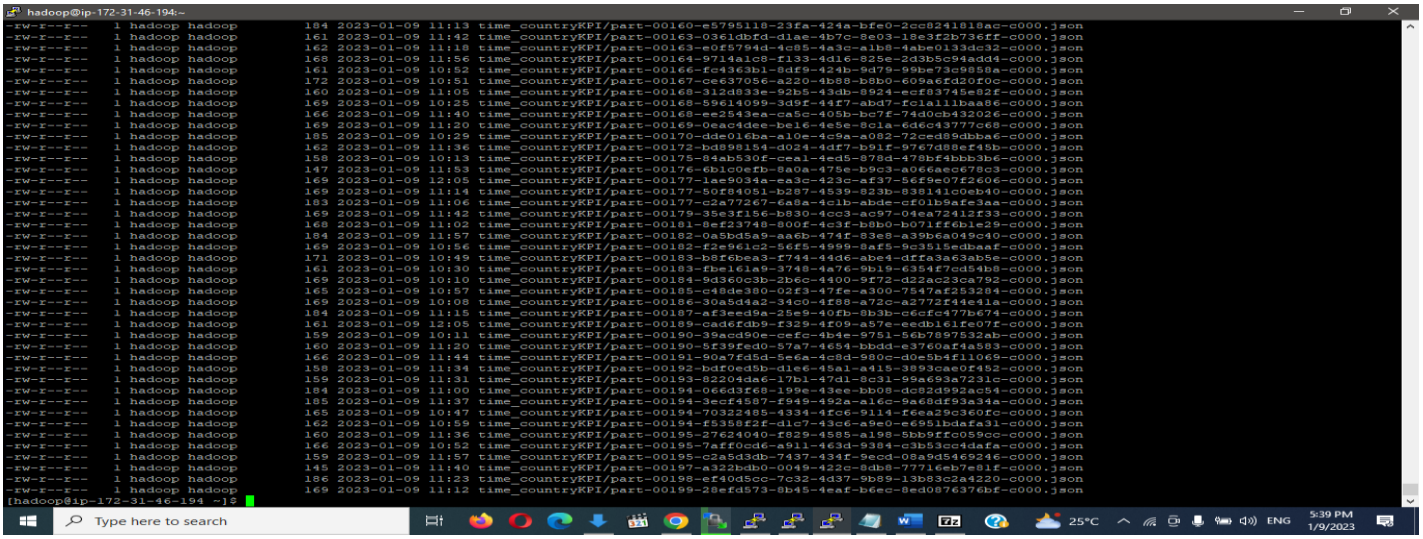
Hdfs dfs -ls timeKPI/



```
-rw-r--r--   1 hadoop hadoop          0 2023-01-09 10:33 timeKPI/part-00000-09cef0fd-38a6-4ff5-8c48-a4251a040c0f-c000.json
-rw-r--r--   1 hadoop hadoop          0 2023-01-09 10:06 timeKPI/part-00000-0a6795e0-a165-49a2-bd8a-85039a097bdf-c000.json
-rw-r--r--   1 hadoop hadoop          0 2023-01-09 10:18 timeKPI/part-00000-0b1dbbb4-d8a8-44dd-b744-76d1f8e31931-c000.json
-rw-r--r--   1 hadoop hadoop          0 2023-01-09 10:23 timeKPI/part-00000-0d9cdeba-ded7-477a-83be-085c23069fef-c000.json
-rw-r--r--   1 hadoop hadoop          0 2023-01-09 10:46 timeKPI/part-00000-0fbd88bd-1c81-43a1-8fd6-521b2403e8bf-c000.json
-rw-r--r--   1 hadoop hadoop          0 2023-01-09 10:08 timeKPI/part-00000-22526a4c-83c5-4f4e-8201-5ab08ed2eed6-c000.json
-rw-r--r--   1 hadoop hadoop          0 2023-01-09 10:22 timeKPI/part-00000-237c9b23-d9de-45e2-92d7-af4b243e07ca-c000.json
-rw-r--r--   1 hadoop hadoop          0 2023-01-09 10:11 timeKPI/part-00000-245c09ab-dbf6-463a-8187-1d8f5bf041f8-c000.json
-rw-r--r--   1 hadoop hadoop          0 2023-01-09 10:41 timeKPI/part-00000-2788d082-5052-4d93-a05e-5d7e1046db29-c000.json
-rw-r--r--   1 hadoop hadoop          0 2023-01-09 10:17 timeKPI/part-00000-29888234-bbdc-42ce-9594-941df98f1313-c000.json
-rw-r--r--   1 hadoop hadoop          0 2023-01-09 10:31 timeKPI/part-00000-2e75a74d-e824-4acf-bf93-a5c258382fb2-c000.json
-rw-r--r--   1 hadoop hadoop          0 2023-01-09 10:25 timeKPI/part-00000-2ec5c456-5c8c-418b-b569-424da567eaa5-c000.json
-rw-r--r--   1 hadoop hadoop          0 2023-01-09 10:29 timeKPI/part-00000-3098b667-4981-4eff-8c29-e6f3b7b1ca80-c000.json
-rw-r--r--   1 hadoop hadoop          0 2023-01-09 10:34 timeKPI/part-00000-3300a794-06ce-49ea-943c-87868ee8a142-c000.json
-rw-r--r--   1 hadoop hadoop          0 2023-01-09 10:48 timeKPI/part-00000-41a7f878-4178-40b8-971b-8e1ca85e2043-c000.json
-rw-r--r--   1 hadoop hadoop          0 2023-01-09 10:05 timeKPI/part-00000-51b84f0a-b273-4d28-a169-29455edeb0f8-c000.json
-rw-r--r--   1 hadoop hadoop          0 2023-01-09 10:20 timeKPI/part-00000-5ed2627b-0064-4192-92f2-5045e21d3732-c000.json
-rw-r--r--   1 hadoop hadoop          0 2023-01-09 10:12 timeKPI/part-00000-5f9643c3-f644-4fe2-84ec-1ac90e41b1b7-c000.json
-rw-r--r--   1 hadoop hadoop          0 2023-01-09 10:42 timeKPI/part-00000-69cc07fa-90ba-499d-a5bd-86c10dd80e36-c000.json
-rw-r--r--   1 hadoop hadoop          0 2023-01-09 10:26 timeKPI/part-00000-6a6f699a-2257-48e2-a949-16fbb6eee85e-c000.json
-rw-r--r--   1 hadoop hadoop          0 2023-01-09 10:38 timeKPI/part-00000-6c55cc3e-5544-4f04-8569-2fa0be32b1e8-c000.json
-rw-r--r--   1 hadoop hadoop          0 2023-01-09 10:45 timeKPI/part-00000-6e6106fc-d1cd-4ecf-b499-bdf6285b2cba-c000.json
-rw-r--r--   1 hadoop hadoop          0 2023-01-09 10:15 timeKPI/part-00000-75205c21-3fe6-41de-a2cc-a6e9b588dad9-c000.json
-rw-r--r--   1 hadoop hadoop          0 2023-01-09 10:37 timeKPI/part-00000-8ac91154-015d-4263-be77-922c7386466c-c000.json
-rw-r--r--   1 hadoop hadoop          0 2023-01-09 10:36 timeKPI/part-00000-8d96e94f-e6d2-4de5-bcf2-9e0b58e3453e-c000.json
-rw-r--r--   1 hadoop hadoop          0 2023-01-09 10:49 timeKPI/part-00000-8e7d04c3-0a74-4c21-a1e9-04ca0c77dfa5-c000.json
-rw-r--r--   1 hadoop hadoop          0 2023-01-09 10:35 timeKPI/part-00000-959a0fbc-6316-4faa-b4aa-da739cb25393-c000.json
-rw-r--r--   1 hadoop hadoop          0 2023-01-09 10:47 timeKPI/part-00000-a295ce3d-6ad2-4518-85dc-b1afc4ba9ff8-c000.json
-rw-r--r--   1 hadoop hadoop          0 2023-01-09 10:25 timeKPI/part-00000-a52658d5-de4e-402c-ac59-42420fb8e373-c000.json
-rw-r--r--   1 hadoop hadoop          0 2023-01-09 10:07 timeKPI/part-00000-a66a4aa7-5b27-4373-a7ae-8ab7c5f6bb22-c000.json
-rw-r--r--   1 hadoop hadoop          0 2023-01-09 10:44 timeKPI/part-00000-afbca605-1c29-4b09-ba77-e487f0353bb3-c000.json
-rw-r--r--   1 hadoop hadoop          0 2023-01-09 10:19 timeKPI/part-00000-b8cfe3ab-f847-4914-832b-7b38f5e4d802-c000.json
-rw-r--r--   1 hadoop hadoop          0 2023-01-09 10:32 timeKPI/part-00000-ba86f0fe-f198-45d9-81db-e8dfa9f67105-c000.json
-rw-r--r--   1 hadoop hadoop          0 2023-01-09 10:14 timeKPI/part-00000-bc362cf7-1a2f-48af-bdb8-20c878835842-c000.json
-rw-r--r--   1 hadoop hadoop          0 2023-01-09 10:13 timeKPI/part-00000-c9b4acd2-4e33-4e18-9915-5e2b0a3aca93-c000.json
-rw-r--r--   1 hadoop hadoop          0 2023-01-09 10:50 timeKPI/part-00000-c9ed12d7-3c51-48db-9f5e-0c2b1967907d-c000.json
-rw-r--r--   1 hadoop hadoop          0 2023-01-09 10:39 timeKPI/part-00000-cc26754c-248a-40e2-aa44-d2251b0c3aa7-c000.json
-rw-r--r--   1 hadoop hadoop          0 2023-01-09 10:27 timeKPI/part-00000-cc33f219-52e3-4364-850d-c1552c8e7a30-c000.json
-rw-r--r--   1 hadoop hadoop          0 2023-01-09 10:43 timeKPI/part-00000-d2624f63-a023-4a87-b2dd-5c2199aaff76-c000.json
-rw-r--r--   1 hadoop hadoop          0 2023-01-09 10:30 timeKPI/part-00000-d435f4b3-73d4-4897-a0b8-6529d81a81b0-c000.json
-rw-r--r--   1 hadoop hadoop          0 2023-01-09 10:09 timeKPI/part-00000-d5f52f25-0cfa-4cfe-b162-4b3e577a9206-c000.json
-rw-r--r--   1 hadoop hadoop          0 2023-01-09 10:16 timeKPI/part-00000-d9ce605b-238b-4ee6-9b10-76e407185bf6-c000.json
-rw-r--r--   1 hadoop hadoop          0 2023-01-09 10:28 timeKPI/part-00000-ec9ef8ff-4d42-41d8-82dc-fa051a3d68d7-c000.json
-rw-r--r--   1 hadoop hadoop          0 2023-01-09 10:10 timeKPI/part-00000-f405bbd1-d57d-40f2-b6eb-709e6b661cc0-c000.json
```

Timekpi.json

Hdfs dfs -ls countrytimeKPI/



```
-rw-r--r--   1 hadoop hadoop        184 2023-01-09 11:13 time_countryKPI/part-00160-e5795118-23fa-424a-bfe0-2cc8241818ac-c000.json
-rw-r--r--   1 hadoop hadoop        161 2023-01-09 11:42 time_countryKPI/part-00163-0361dbfd-d1ae-4b7c-8e03-18e3f2b736ff-c000.json
-rw-r--r--   1 hadoop hadoop        162 2023-01-09 11:18 time_countryKPI/part-00163-e0f5794d-4c85-4a3c-a1b8-4abe0133dc32-c000.json
-rw-r--r--   1 hadoop hadoop        168 2023-01-09 11:56 time_countryKPI/part-00164-9714a1c8-f133-4d16-825e-2d3b5c94add4-c000.json
-rw-r--r--   1 hadoop hadoop        161 2023-01-09 10:52 time_countryKPI/part-00166-fc4363b1-8df9-424b-9d79-99be73c9858a-c000.json
-rw-r--r--   1 hadoop hadoop        172 2023-01-09 10:51 time_countryKPI/part-00167-ce637056-a220-4b88-b8b0-609a6fd20f0c-c000.json
-rw-r--r--   1 hadoop hadoop        160 2023-01-09 11:05 time_countryKPI/part-00168-312d833e-92b5-43db-8924-ecf83745e82f-c000.json
-rw-r--r--   1 hadoop hadoop        169 2023-01-09 10:25 time_countryKPI/part-00168-59614099-3d9f-44f7-abd7-fc1a111baa86-c000.json
-rw-r--r--   1 hadoop hadoop        166 2023-01-09 11:40 time_countryKPI/part-00168-ee2543ea-ca5c-405b-bc7f-74d0cb432026-c000.json
-rw-r--r--   1 hadoop hadoop        169 2023-01-09 11:20 time_countryKPI/part-00169-0eac4dee-be16-4e5e-8c1a-6d6c43777c68-c000.json
-rw-r--r--   1 hadoop hadoop        185 2023-01-09 10:29 time_countryKPI/part-00170-dde016ba-a10e-4c9a-a082-72ced89dbba6-c000.json
-rw-r--r--   1 hadoop hadoop        162 2023-01-09 11:36 time_countryKPI/part-00172-bd898154-d024-4df7-b91f-9767d88ef45b-c000.json
-rw-r--r--   1 hadoop hadoop        158 2023-01-09 10:13 time_countryKPI/part-00175-84ab530f-cea1-4ed5-878d-478bf4bbb3b6-c000.json
-rw-r--r--   1 hadoop hadoop        147 2023-01-09 11:53 time_countryKPI/part-00176-6b1c0efb-8a0a-475e-b9c3-a066aec678c3-c000.json
-rw-r--r--   1 hadoop hadoop        169 2023-01-09 12:05 time_countryKPI/part-00177-1ae9034a-ea3c-423c-af37-56f9e07f2606-c000.json
-rw-r--r--   1 hadoop hadoop        169 2023-01-09 11:14 time_countryKPI/part-00177-50f84051-b287-4539-823b-838141c0eb40-c000.json
-rw-r--r--   1 hadoop hadoop        183 2023-01-09 11:06 time_countryKPI/part-00177-c2a77267-6a8a-4c1b-abde-cf01b9afe3aa-c000.json
-rw-r--r--   1 hadoop hadoop        169 2023-01-09 11:42 time_countryKPI/part-00179-35e3f156-b830-4cc3-ac97-04ea72412f33-c000.json
-rw-r--r--   1 hadoop hadoop        168 2023-01-09 11:02 time_countryKPI/part-00181-8ef23748-800f-4c3f-b8b0-b071ff6b1e29-c000.json
-rw-r--r--   1 hadoop hadoop        184 2023-01-09 11:57 time_countryKPI/part-00182-0a5bd5a9-aa6b-474f-83e8-a39b6a049c40-c000.json
-rw-r--r--   1 hadoop hadoop        169 2023-01-09 10:56 time_countryKPI/part-00182-f2e961c2-56f5-4999-8af5-9c3515edbaaf-c000.json
-rw-r--r--   1 hadoop hadoop        171 2023-01-09 10:49 time_countryKPI/part-00183-b8f6bea3-f744-44d6-abe4-dffa3a63ab5e-c000.json
-rw-r--r--   1 hadoop hadoop        161 2023-01-09 10:30 time_countryKPI/part-00183-fbe161a9-3748-4a76-9b19-6354f7cd54b8-c000.json
-rw-r--r--   1 hadoop hadoop        169 2023-01-09 10:10 time_countryKPI/part-00184-9d360c3b-2b6c-4400-9f72-d22ac23ca792-c000.json
-rw-r--r--   1 hadoop hadoop        165 2023-01-09 10:57 time_countryKPI/part-00185-c48de380-02f3-47fe-a300-7547af253284-c000.json
-rw-r--r--   1 hadoop hadoop        169 2023-01-09 10:08 time_countryKPI/part-00186-30a5d4a2-34c0-4f88-a72c-a2772f44e41a-c000.json
-rw-r--r--   1 hadoop hadoop        184 2023-01-09 11:15 time_countryKPI/part-00187-af3eed9a-25e9-40fb-8b3b-c6cfc477b674-c000.json
-rw-r--r--   1 hadoop hadoop        161 2023-01-09 12:05 time_countryKPI/part-00189-cad6fdb9-f329-4f09-a57e-eedb161fe07f-c000.json
-rw-r--r--   1 hadoop hadoop        159 2023-01-09 10:11 time_countryKPI/part-00190-39acd90e-cefc-4b4e-9751-56b7897532ab-c000.json
-rw-r--r--   1 hadoop hadoop        160 2023-01-09 11:20 time_countryKPI/part-00190-5f39fed0-57a7-4654-bbdd-e3760af4a583-c000.json
-rw-r--r--   1 hadoop hadoop        166 2023-01-09 11:44 time_countryKPI/part-00191-90a7fd5d-5e6a-4c8d-980c-d0e5b4f11069-c000.json
-rw-r--r--   1 hadoop hadoop        158 2023-01-09 11:34 time_countryKPI/part-00192-bdf0ed5b-d1e6-45a1-a415-3893cae0f452-c000.json
-rw-r--r--   1 hadoop hadoop        159 2023-01-09 11:31 time_countryKPI/part-00193-82204da6-17b1-47d1-8c31-99a693a7231c-c000.json
-rw-r--r--   1 hadoop hadoop        184 2023-01-09 11:00 time_countryKPI/part-00194-066d3f68-199e-43ee-bb08-dc82d992ac54-c000.json
-rw-r--r--   1 hadoop hadoop        185 2023-01-09 11:37 time_countryKPI/part-00194-3ecf4587-f949-492a-a16c-9a68df93a34a-c000.json
-rw-r--r--   1 hadoop hadoop        165 2023-01-09 10:47 time_countryKPI/part-00194-70322485-4334-4fc6-9114-f6ea29c360fc-c000.json
-rw-r--r--   1 hadoop hadoop        162 2023-01-09 10:59 time_countryKPI/part-00194-f5358f2f-d1c7-43c6-a9e0-e6951bdafa31-c000.json
-rw-r--r--   1 hadoop hadoop        160 2023-01-09 11:36 time_countryKPI/part-00195-27624040-f829-4585-a198-5bb9ffc059cc-c000.json
-rw-r--r--   1 hadoop hadoop        166 2023-01-09 10:52 time_countryKPI/part-00195-7aff0cd6-a911-463d-9384-c3b53cc4dafa-c000.json
-rw-r--r--   1 hadoop hadoop        159 2023-01-09 11:57 time_countryKPI/part-00195-c2a5d3db-7437-434f-9ecd-08a9d5469246-c000.json
-rw-r--r--   1 hadoop hadoop        145 2023-01-09 11:40 time_countryKPI/part-00197-a322bdb0-0049-422c-8db8-77716eb7e81f-c000.json
-rw-r--r--   1 hadoop hadoop        186 2023-01-09 11:23 time_countryKPI/part-00198-ef40d5cc-7c32-4d37-9b89-13b83c2a4220-c000.json
-rw-r--r--   1 hadoop hadoop        169 2023-01-09 11:12 time_countryKPI/part-00199-28efd573-8b45-4eaf-b6ec-8ed0876376bf-c000.json
[hadoop@ip-172-31-46-194 ~]$
```

Country_timekpi.json

To copy file in hdfs by using get.

Hadoop fs -get timeKPI/ ~/   [for timekpi file]

Hadoop fs -get countrytimeKPI/ ~/ [for countrykpi file]

To make zip file using zip command.

Zip -r output time country console-output