Code Logic - Retail Data Analysis

Spark Streaming Code Implementation:

1. Import the necessary **PySpark** modules such as **SparkSession** class, all the functions and data types from **pyspark.sql.functions** and **pyspark.sql.types** modules respectively.

```
# Import necessary libraries
from pyspark.sql import SparkSession
from pyspark.sql.functions import *
from pyspark.sql.types import *
```

2. Create a Spark Session which is an entry point to interact with Spark application.

```
# Create a spark session
spark = SparkSession \
        .builder \
        .appName("Retail Data Analysis - Spark Streaming") \
        .getOrCreate()

# Set log level to ERROR
spark.sparkContext.setLogLevel('ERROR')
```

3. Connect to Kafka server and read the retail data from it (raw data is in JSON format), define proper schema or data structure to convert it to structured data frame.

```
# Read Retail Dataset from Kafka topic : real-time-project
retail_data_source = spark  \
             .readStream  \
             .format("kafka")  \
             .option("kafka.bootstrap.servers","18.211.252.152:9092")  \
             .option("subscribe","real-time-project")  \
             .load()

# Defining the Data Structure for source retail data (json format).
retail_datastructure = StructType([
                        StructField("invoice_no", LongType()),
                        StructField("country", StringType()),
                        StructField("timestamp", TimestampType()),
                        StructField("type", StringType()),
                        StructField("items", ArrayType(StructType([
                                                StructField("SKU", StringType()),
                                                StructField("title", StringType()),
                                                StructField("unit_price", FloatType()),
                                                StructField("quantity", IntegerType())
                                            ])))
                            ])

# Converting source data from json format to dataframe (table).
retail_data_formatted =  retail_data_source \
                .select(from_json(col("value").cast("string"), retail_datastructure).alias("retail_data")) \
                .select("retail_data.*")
```

4. Define the utility functions to calculate the following metrics:
   a. **Total Cost:**
      Calculate the total cost of all items present in each invoice. Cost for each items list can be calculated by multiplying quantity with unit price. If the invoice type is

"RETURN" then we convert the total cost to negative value which indicates loss of amount.

```
def calculate_total_cost(items, type):
    '''
        Calculate total cost associated for each invoice.
        total_cost = sum of (unit_price * quantity) for all items.
        If the type is Return then the total cost will be negative which represents amount loss.
    '''
    total_cost = 0
    for item in items:
        total_cost = total_cost + (item[2] * item[3])
    if type == 'RETURN':
        return total_cost * -1
    return total_cost
```

b. Total Items:

Calculate the total quantity for all the items list in an invoice be it an "ORDER" or "RETURN" type.

```
def calculate_total_items(items):
    '''
        Calculate total no. of items present in each invoice (quantity).
    '''
    total_items = 0
    for item in items:
        total_items = total_items + item[3]
    return total_items
```

c. Check "ORDER" type:

This function is to check whether the type is ORDER or not. If it is ORDER then we assign the flag as 1, else 0.

```
def check_order_type(type):
    '''
        To check whether the type is Order or not
            1 - ORDER Type
            0 - Not an Order Type
    '''
    if type == 'ORDER':
        return 1
    return 0
```

d. Check "RETURN" type:

This function is to check whether the type is RETURN or not. If it is RETURN then we assign the flag as 1, else 0.

```
def check_return_type(type):
    '''
        To check whether the type is Return or not
            1 - Return Type
            0 - Not a Return Type
    '''
    if type == 'RETURN':
        return 1
    return 0
```

5. Convert the utility functions to UDFs which can then be used to derive the metrics in further transformations.

```
# Converting utility functions to UDFs
get_total_cost = udf(calculate_total_cost, DoubleType())
get_total_items = udf(calculate_total_items, IntegerType())
get_order_flag = udf(check_order_type, IntegerType())
get_return_flag = udf(check_return_type, IntegerType())
```

6. Derive the attributes/metrics such as total_cost, total_items, is_order and is_return based on the source data using the UDFs defined and select the following columns as final data frame and write the output to the console:
    a. invoice_no
    b. country
    c. timestamp
    d. total_cost
    e. total_items
    f. is_order
    g. is_return

```
# Derive the columns/metrics: total_cost, total_items, is_order, is_return from the source dataframe and select the required columns
retail_data_transformed = retail_data_formatted \
                        .withColumn("total_cost", get_total_cost(retail_data_formatted.items, retail_data_formatted.type)) \
                        .withColumn("total_items", get_total_items(retail_data_formatted.items)) \
                        .withColumn("is_order", get_order_flag(retail_data_formatted.type)) \
                        .withColumn("is_return", get_return_flag(retail_data_formatted.type)) \
                        .select("invoice_no", "country", "timestamp", "total_cost", "total_items", "is_order", "is_return")

# Write the retail transformed data to console
retail_data_sink = retail_data_transformed \
                .writeStream \
                .outputMode("append") \
                .format("console") \
                .option("truncate", "false") \
                .trigger(processingTime="1 minute") \
                .start()
```

7. Calculate the following KPIs based on timestamp with one minute window range and one minute sliding interval:
    a. **Orders Per Minute (OPM)**
    b. **total_sale_volume:** Total amount for each minute (window range) – overall sales
    c. **average_transaction_size:** Average amount per minute – overall sales
    d. **rate_of_return:** Average number of returned items per minute.

Write the Time-based KPI output in HDFS (Path: /user/hadoop/Timebased-KPI) in the form of JSON files.

```
# Calculate time based KPIs: OPM (Orders Per Minute), total_sale_volume, average_transaction_size, rate_of_return.
retail_time_kpi = retail_data_transformed \
                .withWatermark("timestamp","1 minute") \
                .groupBy(window("timestamp", "1 minute", "1 minute")) \
                .agg(count("invoice_no").alias("OPM"),
                     sum("total_cost").alias("total_sale_volume"),
                     avg("total_cost").alias("average_transaction_size"),
                     avg("is_return").alias("rate_of_return")) \
                .select("window",
                        "OPM",
                        "total_sale_volume",
                        "average_transaction_size",
                        "rate_of_return")

# Write retail data time based KPI in HDFS (JSON format)
retail_time_kpi_sink = retail_time_kpi \
                     .writeStream \
                     .format("json") \
                     .outputMode("append") \
                     .option("truncate", "false") \
                     .option("path", "Timebased-KPI") \
                     .option("checkpointLocation", "Timebased-Checkpoint") \
                     .trigger(processingTime="1 minute") \
                     .start()
```

8. Calculate the same above KPIs except average_transaction_size but based on time and country wise with one minute window range and one minute sliding interval. Write Time and Country based KPI output in HDFS (Path: /user/hadoop/Country-and-timebased-KPI) in the form of JSON files.

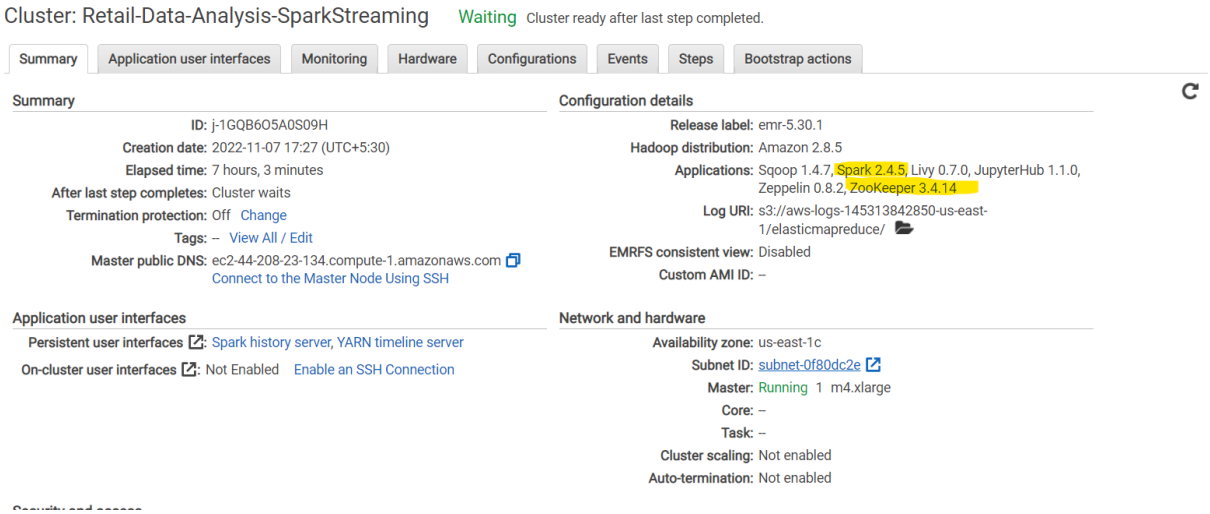
```
# Calculate time-country based KPIs: OPM (Orders Per Minute), total_sale_volume, rate_of_return.
retail_time_country_kpi = retail_data_transformed \
                .withWatermark("timestamp","1 minute") \
                .groupBy(window("timestamp", "1 minute", "1 minute"), "country") \
                .agg(count("invoice_no").alias("OPM"),
                     sum("total_cost").alias("total_sale_volume"),
                     avg("is_return").alias("rate_of_return")) \
                .select("window",
                        "country",
                        "OPM",
                        "total_sale_volume",
                        "rate_of_return")

# Write retail data time and country based KPI in HDFS (JSON format)
retail_time_country_kpi_sink = retail_time_country_kpi \
                     .writeStream \
                     .format("json") \
                     .outputMode("append") \
                     .option("truncate", "false") \
                     .option("path", "Country-and-timebased-KPI") \
                     .option("checkpointLocation", "Country-and-timebased-Checkpoint") \
                     .trigger(processingTime="1 minute") \
                     .start()
```

9. Keep all the streams alive until there is an interruption.

```
# Keep all the streams alive until it is terminated
retail_data_sink.awaitTermination()
retail_time_kpi_sink.awaitTermination()
retail_time_country_kpi_sink.awaitTermination()
```

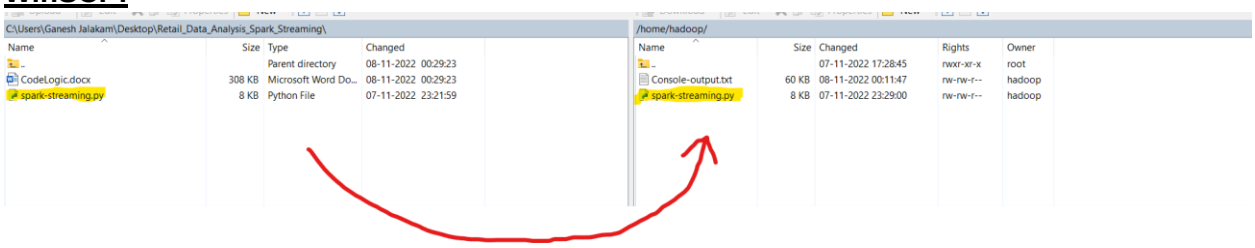
Code Deployment and Execution Steps:

1. Create an EMR Instance with **Spark** and **ZooKeeper** applications installed in it.



2. Copy the python file (spark-streaming.py) under the path: /home/hadoop using WinSCP files transfer application (Windows OS).

WinSCP:



EMR Instance:

```
[hadoop@ip-172-31-82-226 ~]$ ls -ltr *.py
-rw-rw-r-- 1 hadoop hadoop 7374 Nov  7 17:59 spark-streaming.py
[hadoop@ip-172-31-82-226 ~]$ pwd
/home/hadoop
[hadoop@ip-172-31-82-226 ~]$
```

3. Run the following command to enable Kafka Integration with Apache Spark.

   export SPARK_KAFKA_VERSION=0.10

4. Execute the python file using spark-submit command providing Kafka jar package as an argument. Save the console output in a text file (Console-output.txt).

   spark-submit --packages org.apache.spark:spark-sql-kafka-0-10_2.11:2.4.5 spark-streaming.py > Console-output.txt

5. Read the console output file using the command: cat Console-output.txt and check whether we can see the transformed data as per our requirement.

```
+----------+-------+---------+----------+-----------+--------+---------+

-------------------------------------------
Batch: 1
-------------------------------------------
+---------------+--------------+-------------------+------------------+-----------+--------+---------+
|invoice_no     |country       |timestamp          |total_cost        |total_items|is_order|is_return|
+---------------+--------------+-------------------+------------------+-----------+--------+---------+
|154132551996985|United Kingdom|2022-11-07 19:19:03|0.85000002384185791|1          |1       |0        |
|154132551996986|United Kingdom|2022-11-07 19:19:05|35.40000057220459 |12         |1       |0        |
|154132551996987|United Kingdom|2022-11-07 19:19:15|106.2000002861023 |60         |1       |0        |
|154132551996988|United Kingdom|2022-11-07 19:19:31|4.199999809265137 |2          |1       |0        |
|154132551996989|United Kingdom|2022-11-07 19:19:31|44.78000116348266626|1          |1       |0        |
|154132551996990|United Kingdom|2022-11-07 19:19:31|4.130000114440918 |1          |1       |0        |
|154132551996991|United Kingdom|2022-11-07 19:19:40|43.52000045776367 |16         |1       |0        |
|154132551996992|United Kingdom|2022-11-07 19:19:42|13.000000178813934|24         |1       |0        |
|154132551996993|United Kingdom|2022-11-07 19:19:46|256.8399959206581 |65         |1       |0        |
|154132551996994|United Kingdom|2022-11-07 19:19:48|37.899999380111694|22         |1       |0        |
|154132551996995|United Kingdom|2022-11-07 19:19:51|25.0              |2          |1       |0        |
|154132551996996|United Kingdom|2022-11-07 19:19:57|40.63999992609024 |38         |1       |0        |
+---------------+--------------+-------------------+------------------+-----------+--------+---------+

-------------------------------------------
Batch: 2
-------------------------------------------
+---------------+--------------+-------------------+------------------+-----------+--------+---------+
|invoice_no     |country       |timestamp          |total_cost        |total_items|is_order|is_return|
+---------------+--------------+-------------------+------------------+-----------+--------+---------+
|154132551996997|United Kingdom|2022-11-07 19:19:59|60.79999828338623 |14         |1       |0        |
|154132551996998|United Kingdom|2022-11-07 19:20:00|121.480001807212832|52         |1       |0        |
|154132551996999|United Kingdom|2022-11-07 19:20:02|5.879999935626984 |8          |1       |0        |
|154132551997000|United Kingdom|2022-11-07 19:20:06|66.0999984741211  |21         |1       |0        |
|154132551997001|United Kingdom|2022-11-07 19:20:15|34.550000190734863|11         |1       |0        |
|154132551997002|France        |2022-11-07 19:20:16|222.599994659423832|28         |1       |0        |
|154132551997003|United Kingdom|2022-11-07 19:20:29|48.910000920295715|9          |1       |0        |
|154132551997004|United Kingdom|2022-11-07 19:20:30|1.649999976158142 |1          |1       |0        |
|154132551997005|Belgium       |2022-11-07 19:20:33|1.659999966621399 |1          |1       |0        |
|154132551997006|United Kingdom|2022-11-07 19:20:37|9.990000009536743 |7          |1       |0        |
|154132551997007|United Kingdom|2022-11-07 19:20:44|10.809999823570251|13         |1       |0        |
|154132551997008|France        |2022-11-07 19:20:44|73.49999916553497 |30         |1       |0        |
|154132551997009|United Kingdom|2022-11-07 19:20:48|5099.489902853966 |533        |1       |0        |
|154132551997010|Sweden        |2022-11-07 19:20:54|82.00000143051147 |124        |1       |0        |
+---------------+--------------+-------------------+------------------+-----------+--------+---------+

-------------------------------------------
Batch: 3
-------------------------------------------
+---------------+--------------+-------------------+------------------+-----------+--------+---------+
```

6. Now check whether all the JSON files are created for Time based and Time-and-Country based KPIs in HDFS (Path: /user/hadoop/).

```
[hadoop@ip-172-31-82-226 ~]$ hadoop fs -ls /user/hadoop/
Found 5 items
drwxr-xr-x   - hadoop hadoop          0 2022-11-07 19:35 /user/hadoop/.sparkStaging
drwxr-xr-x   - hadoop hadoop          0 2022-11-07 18:05 /user/hadoop/Country-and-timebased-Checkpoint
drwxr-xr-x   - hadoop hadoop          0 2022-11-07 19:35 /user/hadoop/Country-and-timebased-KPI
drwxr-xr-x   - hadoop hadoop          0 2022-11-07 18:05 /user/hadoop/Timebased-Checkpoint
drwxr-xr-x   - hadoop hadoop          0 2022-11-07 19:35 /user/hadoop/Timebased-KPI
[hadoop@ip-172-31-82-226 ~]$
```

Read Timebased-KPI JSON files:

hadoop fs -ls /user/hadoop/Timebased-KPI

```
-rw-r--r--   1 hadoop hadoop        194 2022-11-07 18:26 /user/hadoop/Timebased-KPI/part-00114-e5ea6430-20d1-4b7d-a343-c3aafa306c02-c000.json
-rw-r--r--   1 hadoop hadoop        209 2022-11-07 18:15 /user/hadoop/Timebased-KPI/part-00115-ad85fec7-df7e-4321-b399-51b2d2af6f51-c000.json
-rw-r--r--   1 hadoop hadoop        193 2022-11-07 19:26 /user/hadoop/Timebased-KPI/part-00117-25bfaf90-8042-4d97-a1fa-5fa3e36e0252-c000.json
-rw-r--r--   1 hadoop hadoop        193 2022-11-07 19:26 /user/hadoop/Timebased-KPI/part-00120-3a9d438a-30a3-450c-bf89-f7a366275c2a-c000.json
-rw-r--r--   1 hadoop hadoop        193 2022-11-07 19:26 /user/hadoop/Timebased-KPI/part-00122-0a79d6a6-a1f9-4fb7-9835-3fcfcf7fa721-c000.json
-rw-r--r--   1 hadoop hadoop        210 2022-11-07 19:26 /user/hadoop/Timebased-KPI/part-00124-e02b901f-e7b4-479d-b677-3121222f8493-c000.json
-rw-r--r--   1 hadoop hadoop        193 2022-11-07 18:37 /user/hadoop/Timebased-KPI/part-00126-8edad221-a391-4250-8544-eec66f25e604-c000.json
-rw-r--r--   1 hadoop hadoop        194 2022-11-07 19:26 /user/hadoop/Timebased-KPI/part-00127-4df11b7b-6071-4e44-928b-12d0044b6080-c000.json
-rw-r--r--   1 hadoop hadoop        196 2022-11-07 19:26 /user/hadoop/Timebased-KPI/part-00128-bfbf0ffc-06d3-4b3c-bf23-0b888d08c224-c000.json
-rw-r--r--   1 hadoop hadoop        194 2022-11-07 19:33 /user/hadoop/Timebased-KPI/part-00129-db1d51a8-ae77-443f-99b3-1590b0a531f2-c000.json
-rw-r--r--   1 hadoop hadoop        209 2022-11-07 19:26 /user/hadoop/Timebased-KPI/part-00135-bdefe433-70d9-4576-8862-adc63ef8a75a-c000.json
-rw-r--r--   1 hadoop hadoop        212 2022-11-07 18:18 /user/hadoop/Timebased-KPI/part-00139-220fd44f-22fb-4716-af21-437bb88ce6a4-c000.json
-rw-r--r--   1 hadoop hadoop        210 2022-11-07 19:26 /user/hadoop/Timebased-KPI/part-00139-6e7ee3ce-186e-4820-adde-65c8c12647a3-c000.json
-rw-r--r--   1 hadoop hadoop        196 2022-11-07 18:27 /user/hadoop/Timebased-KPI/part-00140-7cebe36c-c4a6-4207-9849-d4651b021751-c000.json
-rw-r--r--   1 hadoop hadoop        194 2022-11-07 19:26 /user/hadoop/Timebased-KPI/part-00142-57e8a1e8-ec97-4188-9bfd-d7020a0c78f7-c000.json
-rw-r--r--   1 hadoop hadoop        194 2022-11-07 19:26 /user/hadoop/Timebased-KPI/part-00146-b801c336-62dd-42ac-aa6b-fa45f6aa2a9b-c000.json
-rw-r--r--   1 hadoop hadoop        210 2022-11-07 18:22 /user/hadoop/Timebased-KPI/part-00148-4cd0368a-4329-48bf-80db-54aa8fb42629-c000.json
-rw-r--r--   1 hadoop hadoop        210 2022-11-07 19:27 /user/hadoop/Timebased-KPI/part-00149-dcf598fa-7b96-4bb5-a98a-f54945f46b47-c000.json
-rw-r--r--   1 hadoop hadoop        195 2022-11-07 19:26 /user/hadoop/Timebased-KPI/part-00150-5a019ba1-93a4-4ada-9379-aeb31363fb9e-c000.json
-rw-r--r--   1 hadoop hadoop        197 2022-11-07 18:39 /user/hadoop/Timebased-KPI/part-00151-8c072788-ed17-4249-a8b8-0aa47ea569a9-c000.json
-rw-r--r--   1 hadoop hadoop        194 2022-11-07 19:26 /user/hadoop/Timebased-KPI/part-00152-df9746de-66fb-4116-9284-c025e90b94b4-c000.json
-rw-r--r--   1 hadoop hadoop        194 2022-11-07 19:26 /user/hadoop/Timebased-KPI/part-00153-c4335a99-80cc-4995-b331-29d7a8b65984-c000.json
-rw-r--r--   1 hadoop hadoop        194 2022-11-07 18:34 /user/hadoop/Timebased-KPI/part-00155-59bf4475-b821-48ed-9e7f-e2c3b2ff12f1-c000.json
-rw-r--r--   1 hadoop hadoop        194 2022-11-07 19:26 /user/hadoop/Timebased-KPI/part-00156-3c658fcc-9425-4ec6-9820-8cba0d86f473-c000.json
-rw-r--r--   1 hadoop hadoop        194 2022-11-07 18:40 /user/hadoop/Timebased-KPI/part-00160-e97f4991-a574-4d74-819c-fd0be799bcec-c000.json
-rw-r--r--   1 hadoop hadoop        193 2022-11-07 18:41 /user/hadoop/Timebased-KPI/part-00163-f37bc758-a00c-482a-be53-1fc49bd7d8ee-c000.json
-rw-r--r--   1 hadoop hadoop        211 2022-11-07 19:26 /user/hadoop/Timebased-KPI/part-00164-5da7baef-84da-480f-b85e-acca5b872f97-c000.json
-rw-r--r--   1 hadoop hadoop        405 2022-11-07 19:26 /user/hadoop/Timebased-KPI/part-00165-b009cae9-4633-4033-9b02-1b3bf7c1a2fb-c000.json
-rw-r--r--   1 hadoop hadoop        194 2022-11-07 19:30 /user/hadoop/Timebased-KPI/part-00166-871a26c9-976e-43d4-a1b2-dd5ad0ff7bac-c000.json
-rw-r--r--   1 hadoop hadoop        211 2022-11-07 18:14 /user/hadoop/Timebased-KPI/part-00168-ee26c369-7fb5-4dff-aa6c-d3d07e67645d-c000.json
-rw-r--r--   1 hadoop hadoop        194 2022-11-07 19:26 /user/hadoop/Timebased-KPI/part-00169-3c3b7a4f-1ff5-476a-8e45-ae03506b4056-c000.json
-rw-r--r--   1 hadoop hadoop        212 2022-11-07 18:08 /user/hadoop/Timebased-KPI/part-00170-44b18051-14f1-4b10-bc01-f0f169dd9e60-c000.json
-rw-r--r--   1 hadoop hadoop        194 2022-11-07 19:26 /user/hadoop/Timebased-KPI/part-00170-526ce1bd-7818-4078-af2d-9fe757546cd5-c000.json
-rw-r--r--   1 hadoop hadoop        210 2022-11-07 18:09 /user/hadoop/Timebased-KPI/part-00173-c83b299e-e569-45c3-8c56-b49721c5bc75-c000.json
-rw-r--r--   1 hadoop hadoop        211 2022-11-07 18:11 /user/hadoop/Timebased-KPI/part-00174-73d0fcfc-d9b2-4cbd-a07a-f9afe74a2919-c000.json
-rw-r--r--   1 hadoop hadoop        210 2022-11-07 18:11 /user/hadoop/Timebased-KPI/part-00175-8d91dd6d-69eb-4b46-82c1-15908323145c-c000.json
-rw-r--r--   1 hadoop hadoop        194 2022-11-07 19:26 /user/hadoop/Timebased-KPI/part-00178-6dc3c044-ff63-4796-aa15-04c9b9e5f329-c000.json
-rw-r--r--   1 hadoop hadoop        194 2022-11-07 19:26 /user/hadoop/Timebased-KPI/part-00180-acc9f1d1-8f79-404d-9176-f38155699c06-c000.json
-rw-r--r--   1 hadoop hadoop        194 2022-11-07 18:38 /user/hadoop/Timebased-KPI/part-00185-fb7f7367-f57b-4730-87d6-f9cc84f27bb2-c000.json
-rw-r--r--   1 hadoop hadoop        210 2022-11-07 18:18 /user/hadoop/Timebased-KPI/part-00187-12501d46-81cb-49df-abd9-84e20f2179e5-c000.json
-rw-r--r--   1 hadoop hadoop        194 2022-11-07 18:30 /user/hadoop/Timebased-KPI/part-00189-c2b20298-215f-4daa-a10d-6e3bf52e4f6c-c000.json
-rw-r--r--   1 hadoop hadoop        194 2022-11-07 19:26 /user/hadoop/Timebased-KPI/part-00189-ebf49140-4f97-477c-84e8-016076e2e9bf-c000.json
-rw-r--r--   1 hadoop hadoop        194 2022-11-07 19:26 /user/hadoop/Timebased-KPI/part-00190-ac94212c-5d26-423f-9ff6-ae3342daab50-c000.json
-rw-r--r--   1 hadoop hadoop        196 2022-11-07 18:29 /user/hadoop/Timebased-KPI/part-00191-b59b4e63-3d79-4650-ab24-603fc38c6724-c000.json
-rw-r--r--   1 hadoop hadoop        193 2022-11-07 18:35 /user/hadoop/Timebased-KPI/part-00199-74d6d0d3-bd84-40df-9133-c1aa39a30a8d-c000.json
```

hadoop fs -cat /user/hadoop/Timebased-KPI/part*


```
{"window":{"start":"2022-11-07T18:24:00.000Z","end":"2022-11-07T18:25:00.000Z"},"OPM":6,"total_sale_volume":169.0399973988533,"average_transaction_size":28.173332899808884,"rate_of_return":
0.0}
{"window":{"start":"2022-11-07T19:00:00.000Z","end":"2022-11-07T19:01:00.000Z"},"OPM":5,"total_sale_volume":978.3600056171417,"average_transaction_size":195.67200112342834,"rate_of_return":
0.0}
{"window":{"start":"2022-11-07T18:30:00.000Z","end":"2022-11-07T18:31:00.000Z"},"OPM":13,"total_sale_volume":859.5000116229057,"average_transaction_size":66.11538550945428,"rate_of_return":
0.0}
{"window":{"start":"2022-11-07T18:31:00.000Z","end":"2022-11-07T18:32:00.000Z"},"OPM":5,"total_sale_volume":729.5399057412115,"average_transaction_size":145.9079981148243,"rate_of_return":0
.0}
{"window":{"start":"2022-11-07T18:39:00.000Z","end":"2022-11-07T18:40:00.000Z"},"OPM":13,"total_sale_volume":42.43999069929123,"average_transaction_size":3.2646146691762485,"rate_of_return"
:0.07692307692307693}
{"window":{"start":"2022-11-07T18:52:00.000Z","end":"2022-11-07T18:53:00.000Z"},"OPM":9,"total_sale_volume":1015.339966326952,"average_transaction_size":112.81555181410577,"rate_of_return":
0.1111111111111111}
{"window":{"start":"2022-11-07T18:55:00.000Z","end":"2022-11-07T18:56:00.000Z"},"OPM":10,"total_sale_volume":168.06999823451042,"average_transaction_size":16.806999823451044,"rate_of_return
":0.1}
{"window":{"start":"2022-11-07T19:20:00.000Z","end":"2022-11-07T19:21:00.000Z"},"OPM":13,"total_sale_volume":5778.619899213314,"average_transaction_size":444.5092230164088,"rate_of_return":
0.0}
{"window":{"start":"2022-11-07T18:04:00.000Z","end":"2022-11-07T18:05:00.000Z"},"OPM":12,"total_sale_volume":187.05999130010605,"average_transaction_size":15.58833260834217,"rate_of_return"
:0.16666666666666666}
{"window":{"start":"2022-11-07T19:02:00.000Z","end":"2022-11-07T19:03:00.000Z"},"OPM":11,"total_sale_volume":638.1799947321415,"average_transaction_size":58.016363157467741,"rate_of_return":
0.0}
{"window":{"start":"2022-11-07T17:58:00.000Z","end":"2022-11-07T17:59:00.000Z"},"OPM":12,"total_sale_volume":286.32999593019485,"average_transaction_size":23.860832994182903,"rate_of_return
":0.08333333333333333}
{"window":{"start":"2022-11-07T19:08:00.000Z","end":"2022-11-07T19:09:00.000Z"},"OPM":5,"total_sale_volume":338.95000195503235,"average_transaction_size":67.79000039100647,"rate_of_return":
0.0}
{"window":{"start":"2022-11-07T17:59:00.000Z","end":"2022-11-07T18:00:00.000Z"},"OPM":15,"total_sale_volume":923.0799910724163,"average_transaction_size":61.53866607149442,"rate_of_return":
0.06666666666666667}
{"window":{"start":"2022-11-07T18:02:00.000Z","end":"2022-11-07T18:03:00.000Z"},"OPM":14,"total_sale_volume":1095.5799964368343,"average_transaction_size":78.25571403120246,"rate_of_return"
:0.07142857142857142}
{"window":{"start":"2022-11-07T18:01:00.000Z","end":"2022-11-07T18:02:00.000Z"},"OPM":12,"total_sale_volume":889.1900013685226,"average_transaction_size":74.09916678071022,"rate_of_return":
0.08333333333333333}
{"window":{"start":"2022-11-07T18:46:00.000Z","end":"2022-11-07T18:47:00.000Z"},"OPM":10,"total_sale_volume":532.0199964940548,"average_transaction_size":53.20199964940548,"rate_of_return":
0.0}
{"window":{"start":"2022-11-07T18:53:00.000Z","end":"2022-11-07T18:54:00.000Z"},"OPM":10,"total_sale_volume":507.8799990415573,"average_transaction_size":50.78799990415573,"rate_of_return":
0.0}
{"window":{"start":"2022-11-07T18:28:00.000Z","end":"2022-11-07T18:29:00.000Z"},"OPM":9,"total_sale_volume":521.0200008153915,"average_transaction_size":57.891111201710174,"rate_of_return":
0.0}
{"window":{"start":"2022-11-07T18:08:00.000Z","end":"2022-11-07T18:09:00.000Z"},"OPM":12,"total_sale_volume":458.82999137043953,"average_transaction_size":38.2358326142033,"rate_of_return":
0.08333333333333333}
{"window":{"start":"2022-11-07T18:20:00.000Z","end":"2022-11-07T18:21:00.000Z"},"OPM":8,"total_sale_volume":976.8099947422743,"average_transaction_size":122.10124934278429,"rate_of_return":
0.0}
{"window":{"start":"2022-11-07T18:45:00.000Z","end":"2022-11-07T18:46:00.000Z"},"OPM":8,"total_sale_volume":1014.2899831533432,"average_transaction_size":126.7862478941679,"rate_of_return":
0.0}
{"window":{"start":"2022-11-07T19:01:00.000Z","end":"2022-11-07T19:02:00.000Z"},"OPM":13,"total_sale_volume":980.4299919307232,"average_transaction_size":75.41769168697871,"rate_of_return":
0.0}
{"window":{"start":"2022-11-07T18:19:00.000Z","end":"2022-11-07T18:20:00.000Z"},"OPM":8,"total_sale_volume":340.1900028884411,"average_transaction_size":42.523750361055136,"rate_of_return":
0.125}
{"window":{"start":"2022-11-07T18:25:00.000Z","end":"2022-11-07T18:26:00.000Z"},"OPM":8,"total_sale_volume":1730.93994140625,"average_transaction_size":216.36749267578125,"rate_of_return":0
.0}
```


Read Country-and-timebased-KPI JSON files:

hadoop fs -ls /user/hadoop/Country-and-timebased-KPI

```
-rw-r--r--   1 hadoop hadoop        167 2022-11-07 19:26 /user/hadoop/Country-and-timebased-KPI/part-00130-f8a04585-058b-4f2b-87bc-750907bd944d-c000.json
-rw-r--r--   1 hadoop hadoop        351 2022-11-07 19:26 /user/hadoop/Country-and-timebased-KPI/part-00132-8bb8b272-23da-4e0b-94a9-b7cab11516ba-c000.json
-rw-r--r--   1 hadoop hadoop        167 2022-11-07 19:26 /user/hadoop/Country-and-timebased-KPI/part-00133-566e0488-9b03-4a91-891d-6cdbd359cd5a-c000.json
-rw-r--r--   1 hadoop hadoop        176 2022-11-07 18:31 /user/hadoop/Country-and-timebased-KPI/part-00134-c899c8ee-b12f-4433-a256-d7ebc14c5f7d-c000.json
-rw-r--r--   1 hadoop hadoop        170 2022-11-07 18:18 /user/hadoop/Country-and-timebased-KPI/part-00135-8bc841db-786a-4782-ac4b-ac06afd17d14-c000.json
-rw-r--r--   1 hadoop hadoop        523 2022-11-07 19:26 /user/hadoop/Country-and-timebased-KPI/part-00138-6d2ba6f6-b6ca-44dd-9ec2-a1ead40eeeff-c000.json
-rw-r--r--   1 hadoop hadoop        168 2022-11-07 19:32 /user/hadoop/Country-and-timebased-KPI/part-00140-f529f6a4-44bb-4789-9b27-46e8c52d9a3a-c000.json
-rw-r--r--   1 hadoop hadoop        176 2022-11-07 19:26 /user/hadoop/Country-and-timebased-KPI/part-00141-cf5e0e01-706a-477b-9561-7b131daddadc-c000.json
-rw-r--r--   1 hadoop hadoop        155 2022-11-07 19:26 /user/hadoop/Country-and-timebased-KPI/part-00142-bb4c5083-bb37-407c-ba7a-7bcdc61c8aba-c000.json
-rw-r--r--   1 hadoop hadoop        175 2022-11-07 19:26 /user/hadoop/Country-and-timebased-KPI/part-00143-91e2414f-1ed6-4d2d-8f77-7cdd0e18e169-c000.json
-rw-r--r--   1 hadoop hadoop        175 2022-11-07 19:26 /user/hadoop/Country-and-timebased-KPI/part-00144-0875251e-b55f-496f-8a79-a50449d15158-c000.json
-rw-r--r--   1 hadoop hadoop        176 2022-11-07 19:26 /user/hadoop/Country-and-timebased-KPI/part-00146-46f975b1-1f05-4390-925c-06ed1cdec6aa-c000.json
-rw-r--r--   1 hadoop hadoop        176 2022-11-07 18:10 /user/hadoop/Country-and-timebased-KPI/part-00148-7b3dc2c3-a7b8-4c50-8542-c785f9055d3b-c000.json
-rw-r--r--   1 hadoop hadoop        176 2022-11-07 19:26 /user/hadoop/Country-and-timebased-KPI/part-00152-1e228b96-b8c8-4416-90fe-4cbe84bd5741-c000.json
-rw-r--r--   1 hadoop hadoop        152 2022-11-07 18:21 /user/hadoop/Country-and-timebased-KPI/part-00152-451e253e-8e76-4296-ba8e-f8c49fff3279-c000.json
-rw-r--r--   1 hadoop hadoop        176 2022-11-07 18:34 /user/hadoop/Country-and-timebased-KPI/part-00152-7c5b7d6b-c51a-4817-9b5b-a4e3a71862c1-c000.json
-rw-r--r--   1 hadoop hadoop        176 2022-11-07 18:26 /user/hadoop/Country-and-timebased-KPI/part-00153-61e8bc5a-1c4d-48c8-aa5c-96399fabaf26-c000.json
-rw-r--r--   1 hadoop hadoop        175 2022-11-07 19:26 /user/hadoop/Country-and-timebased-KPI/part-00153-aa6c0012-7e8b-40c8-9966-fda5c9e16ada-c000.json
-rw-r--r--   1 hadoop hadoop        347 2022-11-07 19:26 /user/hadoop/Country-and-timebased-KPI/part-00154-01b3db4d-4d19-4d77-b770-4cbbdc2dfc0e-c000.json
-rw-r--r--   1 hadoop hadoop        156 2022-11-07 18:31 /user/hadoop/Country-and-timebased-KPI/part-00156-311e95b0-2d10-45c8-a4cf-620c08a3e1f2-c000.json
-rw-r--r--   1 hadoop hadoop        176 2022-11-07 19:26 /user/hadoop/Country-and-timebased-KPI/part-00157-cc74e2a4-c643-47e5-9bc4-6e26aeb674ee-c000.json
-rw-r--r--   1 hadoop hadoop        168 2022-11-07 19:26 /user/hadoop/Country-and-timebased-KPI/part-00159-5aa877ff-0916-456c-a193-46b523207c7e-c000.json
-rw-r--r--   1 hadoop hadoop        165 2022-11-07 19:26 /user/hadoop/Country-and-timebased-KPI/part-00162-ce63dfe0-b859-4ffd-b13d-ce12c213fa3d-c000.json
-rw-r--r--   1 hadoop hadoop        170 2022-11-07 19:26 /user/hadoop/Country-and-timebased-KPI/part-00163-19c429ba-8955-4ab3-af05-af80776136bb-c000.json
-rw-r--r--   1 hadoop hadoop        159 2022-11-07 19:33 /user/hadoop/Country-and-timebased-KPI/part-00164-a2cfaf7d-810a-4186-a991-e500069f6d8d-c000.json
-rw-r--r--   1 hadoop hadoop        168 2022-11-07 18:31 /user/hadoop/Country-and-timebased-KPI/part-00168-c1fd50eb-26f8-4fe3-a0bb-7526ee054913-c000.json
-rw-r--r--   1 hadoop hadoop        343 2022-11-07 19:26 /user/hadoop/Country-and-timebased-KPI/part-00172-355b9923-dab8-437a-b61e-330a928df7c6-c000.json
-rw-r--r--   1 hadoop hadoop        344 2022-11-07 19:26 /user/hadoop/Country-and-timebased-KPI/part-00174-6cbbafd7-7d2a-4d37-919e-507460f78976-c000.json
-rw-r--r--   1 hadoop hadoop        169 2022-11-07 18:34 /user/hadoop/Country-and-timebased-KPI/part-00176-72f9bad6-a081-4af4-8167-98702cb2cb9f-c000.json
-rw-r--r--   1 hadoop hadoop        166 2022-11-07 19:26 /user/hadoop/Country-and-timebased-KPI/part-00176-fe51f1d2-c06e-43ae-81ab-87d345b98d65-c000.json
-rw-r--r--   1 hadoop hadoop        176 2022-11-07 18:08 /user/hadoop/Country-and-timebased-KPI/part-00178-3529cae7-1ebe-44ec-bee9-4c9a11869f30-c000.json
-rw-r--r--   1 hadoop hadoop        177 2022-11-07 19:26 /user/hadoop/Country-and-timebased-KPI/part-00179-0a263d44-ff51-46be-a0f1-fec1b96488dc-c000.json
-rw-r--r--   1 hadoop hadoop        168 2022-11-07 19:30 /user/hadoop/Country-and-timebased-KPI/part-00182-256fe797-76bd-4f80-bd77-b2e809cf53d1-c000.json
-rw-r--r--   1 hadoop hadoop        166 2022-11-07 18:22 /user/hadoop/Country-and-timebased-KPI/part-00182-583e06c6-279c-4021-a3d7-aa3840f56935-c000.json
-rw-r--r--   1 hadoop hadoop        176 2022-11-07 18:28 /user/hadoop/Country-and-timebased-KPI/part-00183-8628f5d3-6499-467d-a370-7db4782385c7-c000.json
-rw-r--r--   1 hadoop hadoop        192 2022-11-07 18:11 /user/hadoop/Country-and-timebased-KPI/part-00183-dddf6ba4-cf51-4744-9c79-4e37f0903ad8-c000.json
-rw-r--r--   1 hadoop hadoop        192 2022-11-07 19:26 /user/hadoop/Country-and-timebased-KPI/part-00184-b96ae036-6e99-482d-a055-9956b620e953-c000.json
-rw-r--r--   1 hadoop hadoop        176 2022-11-07 18:32 /user/hadoop/Country-and-timebased-KPI/part-00187-49406bdd-e04e-46b7-b132-80f0e5d1b75e-c000.json
-rw-r--r--   1 hadoop hadoop        177 2022-11-07 18:18 /user/hadoop/Country-and-timebased-KPI/part-00187-4acf9047-20f4-4d8d-a802-6bb6a1573df8-c000.json
-rw-r--r--   1 hadoop hadoop        176 2022-11-07 19:34 /user/hadoop/Country-and-timebased-KPI/part-00189-2838e3d2-17a1-4241-b177-9de02d947734-c000.json
-rw-r--r--   1 hadoop hadoop        191 2022-11-07 18:15 /user/hadoop/Country-and-timebased-KPI/part-00190-b665a433-40d9-4297-bb63-6693bc8d2c33-c000.json
-rw-r--r--   1 hadoop hadoop        165 2022-11-07 19:26 /user/hadoop/Country-and-timebased-KPI/part-00191-c93f0425-1511-46ff-afb4-47326e167666-c000.json
-rw-r--r--   1 hadoop hadoop        168 2022-11-07 19:26 /user/hadoop/Country-and-timebased-KPI/part-00193-88214fe2-9211-450b-af8d-1b5586b0f1e0-c000.json
-rw-r--r--   1 hadoop hadoop        175 2022-11-07 19:26 /user/hadoop/Country-and-timebased-KPI/part-00194-bf14c51b-4b89-4680-ae46-2a91337ff3d5-c000.json
-rw-r--r--   1 hadoop hadoop        192 2022-11-07 18:14 /user/hadoop/Country-and-timebased-KPI/part-00197-e9d67e0e-4812-4ce9-aa60-d2db3d8d2f16-c000.json
-rw-r--r--   1 hadoop hadoop        175 2022-11-07 19:26 /user/hadoop/Country-and-timebased-KPI/part-00198-5e770d31-e1e3-4e12-a38f-4feedcc007e1-c000.json
-rw-r--r--   1 hadoop hadoop        192 2022-11-07 18:09 /user/hadoop/Country-and-timebased-KPI/part-00198-fb981706-e868-493a-878e-e740cc3e705b-c000.json
-rw-r--r--   1 hadoop hadoop        176 2022-11-07 19:35 /user/hadoop/Country-and-timebased-KPI/part-00199-ded38657-c4fd-44d1-a9c6-21857a72c55c-c000.json
```

hadoop fs -cat /user/hadoop/Country-and-timebased-KPI/part*


```
{"window":{"start":"2022-11-07T18:41:00.000Z","end":"2022-11-07T18:42:00.000Z"},"country":"Germany","OPM":1,"total_sale_volume":45.0,"rate_of_return":0.0}
{"window":{"start":"2022-11-07T18:34:00.000Z","end":"2022-11-07T18:35:00.000Z"},"country":"United Kingdom","OPM":4,"total_sale_volume":603.0699898898602,"rate_of_return":0.0}
{"window":{"start":"2022-11-07T19:06:00.000Z","end":"2022-11-07T19:07:00.000Z"},"country":"United Kingdom","OPM":8,"total_sale_volume":614.4099923670292,"rate_of_return":0.0}
{"window":{"start":"2022-11-07T18:57:00.000Z","end":"2022-11-07T18:58:00.000Z"},"country":"Channel Islands","OPM":1,"total_sale_volume":69.20999908447266,"rate_of_return":0.0}
{"window":{"start":"2022-11-07T18:00:00.000Z","end":"2022-11-07T18:01:00.000Z"},"country":"United Kingdom","OPM":11,"total_sale_volume":549.3799982666969,"rate_of_return":0.0}
{"window":{"start":"2022-11-07T18:46:00.000Z","end":"2022-11-07T18:47:00.000Z"},"country":"United Kingdom","OPM":10,"total_sale_volume":532.0199964940548,"rate_of_return":0.0}
{"window":{"start":"2022-11-07T18:11:00.000Z","end":"2022-11-07T18:12:00.000Z"},"country":"EIRE","OPM":1,"total_sale_volume":3.75,"rate_of_return":0.0}
{"window":{"start":"2022-11-07T18:24:00.000Z","end":"2022-11-07T18:25:00.000Z"},"country":"United Kingdom","OPM":5,"total_sale_volume":138.83999782800674,"rate_of_return":0.0}
{"window":{"start":"2022-11-07T18:16:00.000Z","end":"2022-11-07T18:17:00.000Z"},"country":"United Kingdom","OPM":10,"total_sale_volume":413.6200016438961,"rate_of_return":0.0}
{"window":{"start":"2022-11-07T19:07:00.000Z","end":"2022-11-07T19:08:00.000Z"},"country":"United Kingdom","OPM":6,"total_sale_volume":285.1399952173233,"rate_of_return":0.0}
{"window":{"start":"2022-11-07T19:00:00.000Z","end":"2022-11-07T19:01:00.000Z"},"country":"Switzerland","OPM":1,"total_sale_volume":283.1499999165535,"rate_of_return":0.0}
{"window":{"start":"2022-11-07T19:03:00.000Z","end":"2022-11-07T19:04:00.000Z"},"country":"United Kingdom","OPM":8,"total_sale_volume":557.1100006103516,"rate_of_return":0.0}
{"window":{"start":"2022-11-07T18:21:00.000Z","end":"2022-11-07T18:22:00.000Z"},"country":"Germany","OPM":1,"total_sale_volume":135.0,"rate_of_return":0.0}
{"window":{"start":"2022-11-07T19:08:00.000Z","end":"2022-11-07T19:09:00.000Z"},"country":"United Kingdom","OPM":5,"total_sale_volume":338.9500019550323 5,"rate_of_return":0.0}
{"window":{"start":"2022-11-07T18:51:00.000Z","end":"2022-11-07T18:52:00.000Z"},"country":"France","OPM":1,"total_sale_volume":215.4700021737744,"rate_of_return":0.0}
{"window":{"start":"2022-11-07T18:44:00.000Z","end":"2022-11-07T18:45:00.000Z"},"country":"EIRE","OPM":1,"total_sale_volume":21.28999999165535,"rate_of_return":0.0}
{"window":{"start":"2022-11-07T19:15:00.000Z","end":"2022-11-07T19:16:00.000Z"},"country":"Portugal","OPM":1,"total_sale_volume":35.2399997711181 64,"rate_of_return":0.0}
{"window":{"start":"2022-11-07T19:23:00.000Z","end":"2022-11-07T19:24:00.000Z"},"country":"Netherlands","OPM":1,"total_sale_volume":15.0,"rate_of_return":0.0}
{"window":{"start":"2022-11-07T18:21:00.000Z","end":"2022-11-07T18:22:00.000Z"},"country":"Portugal","OPM":1,"total_sale_volume":134.650000333786,"rate_of_return":0.0}
{"window":{"start":"2022-11-07T19:04:00.000Z","end":"2022-11-07T19:05:00.000Z"},"country":"United Kingdom","OPM":8,"total_sale_volume":292.839999884367,"rate_of_return":0.0}
{"window":{"start":"2022-11-07T18:51:00.000Z","end":"2022-11-07T18:52:00.000Z"},"country":"Germany","OPM":1,"total_sale_volume":26.350000619888306,"rate_of_return":0.0}
{"window":{"start":"2022-11-07T19:16:00.000Z","end":"2022-11-07T19:17:00.000Z"},"country":"Germany","OPM":1,"total_sale_volume":7.149999737739563,"rate_of_return":0.0}
{"window":{"start":"2022-11-07T18:56:00.000Z","end":"2022-11-07T18:57:00.000Z"},"country":"United Kingdom","OPM":2,"total_sale_volume":227.23999667167664,"rate_of_return":0.0}
{"window":{"start":"2022-11-07T18:24:00.000Z","end":"2022-11-07T18:25:00.000Z"},"country":"Germany","OPM":1,"total_sale_volume":30.199999570846558,"rate_of_return":0.0}
{"window":{"start":"2022-11-07T18:56:00.000Z","end":"2022-11-07T18:57:00.000Z"},"country":"EIRE","OPM":1,"total_sale_volume":18.050000548362732,"rate_of_return":0.0}
{"window":{"start":"2022-11-07T17:58:00.000Z","end":"2022-11-07T17:59:00.000Z"},"country":"United Kingdom","OPM":10,"total_sale_volume":202.7699984908104,"rate_of_return":0.1}
{"window":{"start":"2022-11-07T18:52:00.000Z","end":"2022-11-07T18:53:00.000Z"},"country":"United Kingdom","OPM":8,"total_sale_volume":986.5099671185017,"rate_of_return":0.125}
{"window":{"start":"2022-11-07T19:20:00.000Z","end":"2022-11-07T19:21:00.000Z"},"country":"Belgium","OPM":1,"total_sale_volume":1.659999666621399,"rate_of_return":0.0}
{"window":{"start":"2022-11-07T18:12:00.000Z","end":"2022-11-07T18:13:00.000Z"},"country":"EIRE","OPM":1,"total_sale_volume":-7.949999809265137,"rate_of_return":1.0}
{"window":{"start":"2022-11-07T18:18:00.000Z","end":"2022-11-07T18:19:00.000Z"},"country":"United Kingdom","OPM":15,"total_sale_volume":837.1499945223331,"rate_of_return":0.0}
{"window":{"start":"2022-11-07T18:01:00.000Z","end":"2022-11-07T18:02:00.000Z"},"country":"United Kingdom","OPM":12,"total_sale_volume":889.1900013685226,"rate_of_return":0.0833333333333333
3}
{"window":{"start":"2022-11-07T18:59:00.000Z","end":"2022-11-07T19:00:00.000Z"},"country":"United Kingdom","OPM":13,"total_sale_volume":786.2199989557266,"rate_of_return":0.1538461538461538
5}
{"window":{"start":"2022-11-07T18:22:00.000Z","end":"2022-11-07T18:23:00.000Z"},"country":"United Kingdom","OPM":10,"total_sale_volume":645.0700005590916,"rate_of_return":0.1}
{"window":{"start":"2022-11-07T18:09:00.000Z","end":"2022-11-07T18:10:00.000Z"},"country":"United Kingdom","OPM":11,"total_sale_volume":1313.8999728262424,"rate_of_return":0.0}
{"window":{"start":"2022-11-07T19:24:00.000Z","end":"2022-11-07T19:25:00.000Z"},"country":"United Kingdom","OPM":4,"total_sale_volume":381.0399850904941 6,"rate_of_return":0.0}
{"window":{"start":"2022-11-07T18:05:00.000Z","end":"2022-11-07T18:06:00.000Z"},"country":"United Kingdom","OPM":9,"total_sale_volume":445.0699958801269 5,"rate_of_return":0.1111111111111111
}
{"window":{"start":"2022-11-07T18:50:00.000Z","end":"2022-11-07T18:51:00.000Z"},"country":"EIRE","OPM":1,"total_sale_volume":38.60000014305115,"rate_of_return":0.0}
{"window":{"start":"2022-11-07T19:06:00.000Z","end":"2022-11-07T19:07:00.000Z"},"country":"Cyprus","OPM":1,"total_sale_volume":21.269999980926514,"rate_of_return":0.0}
{"window":{"start":"2022-11-07T19:00:00.000Z","end":"2022-11-07T19:01:00.000Z"},"country":"United Kingdom","OPM":3,"total_sale_volume":635.6100043654442,"rate_of_return":0.0}
{"window":{"start":"2022-11-07T18:04:00.000Z","end":"2022-11-07T18:05:00.000Z"},"country":"United Kingdom","OPM":11,"total_sale_volume":186.7699913084507,"rate_of_return":0.1818181818181818
2}
{"window":{"start":"2022-11-07T19:13:00.000Z","end":"2022-11-07T19:14:00.000Z"},"country":"United Kingdom","OPM":9,"total_sale_volume":822.5999949276447,"rate_of_return":0.0}
{"window":{"start":"2022-11-07T17:59:00.000Z","end":"2022-11-07T18:00:00.000Z"},"country":"United Kingdom","OPM":15,"total_sale_volume":923.0799910724163,"rate_of_return":0.0666666666666666
7}
{"window":{"start":"2022-11-07T19:25:00.000Z","end":"2022-11-07T19:26:00.000Z"},"country":"United Kingdom","OPM":13,"total_sale_volume":1340.830010920763,"rate_of_return":0.0}
```

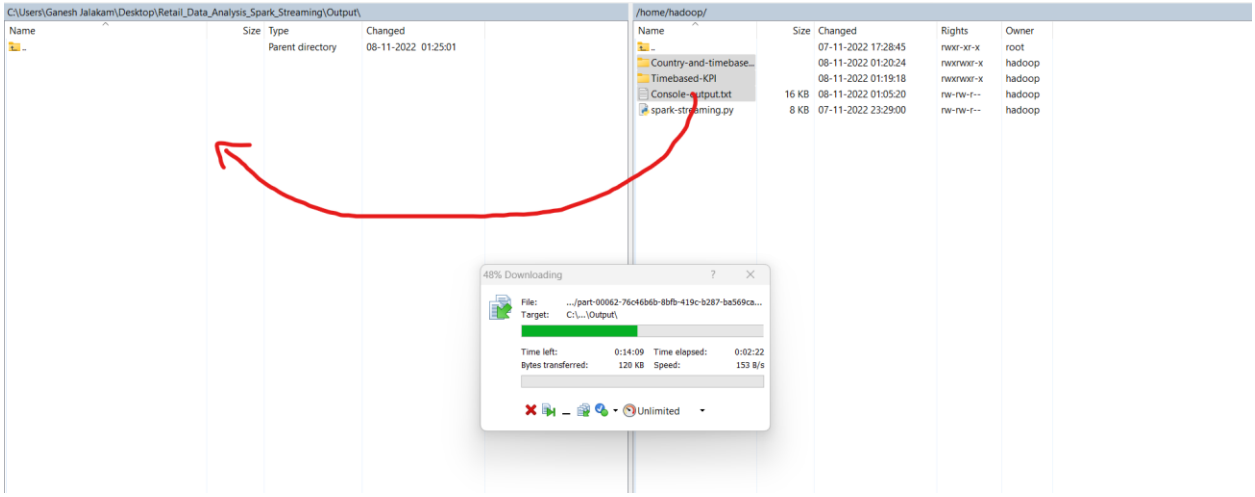
7. Copy Time based and Country-and-Time based KPIs directories from HDFS to local path (/home/hadoop) in EMR instance.

   hadoop fs -get /user/hadoop/Timebased-KPI /home/hadoop/

   hadoop fs -get /user/hadoop/Country-and-timebased-KPI /home/hadoop/

```
[hadoop@ip-172-31-82-226 ~]$ ls -ltr
total 60
-rw-rw-r-- 1 hadoop hadoop  7374 Nov  7 17:59 spark-streaming.py
-rw-rw-r-- 1 hadoop hadoop 16349 Nov  7 19:35 Console-output.txt
drwxrwxr-x 3 hadoop hadoop 12288 Nov  7 19:49 Timebased-KPI
drwxrwxr-x 3 hadoop hadoop 16384 Nov  7 19:50 Country-and-timebased-KPI
[hadoop@ip-172-31-82-226 ~]$
```

8. Copy all the console and JSON output files from EMR to local machine using WinSCP.



9. Post completion of all the above steps, terminate the EMR instance from AWS console.