

# Code Logic - Retail Data Analysis

Logic for Python Script 'spark-streaming.py'

Importing Dependencies :

```
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
```

Writing the Python functions, which contain the logic for the UDFs

is\_a\_order: 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\_a\_return: 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
```

total\_item\_count: 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
```

total\_cost: - 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. The sum of this cost across the products in that invoice gives me the total cost of the order. I also made sure that if the transaction is a return transaction, the total cost is negative.

```
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
```

Initializing the Spark session and setting the log level to error as a good practice

```
spark = SparkSession \  
    .builder \  
    .appName("StructuredSocketRead") \  
    .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 StructField 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.*")
```

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", "total_cost", "total_items", "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

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

Calculating time-and-country-based KPIs (Total sale volume, OPM, Rate of return) having tumbling window of one minute and watermark of one minute. Here I grouped by window and country both

```
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", "country_kpi/") \
    .option("checkpointLocation", "country_kpi/cp/") \
    .trigger(processingTime="1 minutes") \
    .start()
```

Indicating Spark to await termination

```
ByTime_country.awaitTermination()
ByTime.awaitTermination()
order_query_console.awaitTermination()
```

## Console Commands

**The new EMR console will become the default console on Feb 28, 2023**  
[Switch to the new console](#). If you want, you can still switch back. [Learn more](#)

[Clone](#)[Terminate](#)[AWS CLI export](#)

Cluster: My cluster\_Spark **Starting** Configuring cluster software

[Summary](#)[Application user interfaces](#)[Monitoring](#)[Hardware](#)[Configurations](#)[Events](#)[Steps](#)[Bootstrap actions](#)

### Summary

ID: j-JNGH65PW1CID

Creation date: 2023-02-11 13:02 (UTC+5:30)

Elapsed time: 6 minutes

After last step completes: Cluster waits

Termination protection: Off [Change](#)

Tags: -- [View All / Edit](#)

Master public DNS: ec2-3-92-24-30.compute-1.amazonaws.com

[Connect to the Master Node Using SSH](#)

### Configuration details

Release label: emr-5.30.1

Hadoop distribution: Amazon 2.8.5

I started by logging into the ec2 instance as 'hadoop' user.

Setting Environment before running the Spark Job:

```
export SPARK_KAFKA_VERSION=0.10
```

Commands used to run the Spark Submit job:

```
hadoop@ip-172-31-36-187 ~]$ spark-submit --packages org.apache.spark:spark-sql-kafka-0-10_2.11:2.4.5 spark-streaming.py 18.211.252.152 9092 real-time-project
```

```
hadoop@ip-172-31-41-58:~
E:::::EEEEEEEEEE M::::M M::::M M::::M R::RRRRRR:::R
E::::E M::::M M::M M::::M R::R R::::R
E::::E EEEEE M::::M MMM M::::M R::R R::::R
EE:::::EEEEEEEE:::E M::::M M::::M R::R R::::R
E:::::EEEEEEEE:::E M::::M M::::M RR::::R R::::R
EEEEEEEEEEEEEEEEEE MMMMMM MMMMMM RRRRRRR RRRRRR

[hadoop@ip-172-31-41-58 ~]$ ll
total 8
-rw-rw-r-- 1 hadoop hadoop 4464 Feb 11 07:25 spark-streaming.py
[hadoop@ip-172-31-41-58 ~]$ spark-submit --packages org.apache.spark:spark-sql-kafka-0-10_2.11:2.4.5 spark-streaming.py 18.211.252.152 9092 real-time-project
Ivy Default Cache set to: /home/hadoop/.ivy2/cache
The jars for the packages stored in: /home/hadoop/.ivy2/jars
:: loading settings :: url = jar:file:/usr/lib/spark/jars/ivy-2.4.0.jar!/org/apache/ivy/core/settings/ivysettings.xml
org.apache.spark#spark-sql-kafka-0-10_2.11 added as a dependency
:: resolving dependencies :: org.apache.spark#spark-submit-parent-c43060d5-b2fa-4fa2-9342-1b16bfa68620;1.0
  confs: [default]
  found org.apache.spark#spark-sql-kafka-0-10_2.11;2.4.5 in central
  found org.apache.kafka#kafka-clients;2.0.0 in central
  found org.lz4#lz4-java;1.4.0 in central
  found org.xerial.snappy#snappy-java;1.1.7.3 in central
```

```
hadoop@ip-172-31-41-58:~
-2.0.0.jar
23/02/11 07:39:58 INFO spark.SparkContext: Added file file:///home/hadoop/.ivy2/jars/org.spark-project.spark_unused-1.0.0.jar at file:///home/hadoop/.ivy2/jars/org.spark-project.spark_unused-1.0.0.jar with timestamp 1676101198984
23/02/11 07:39:58 INFO util.Utils: Copying /home/hadoop/.ivy2/jars/org.spark-project.spark_unused-1.0.0.jar to /mnt/tmp/spark-3f8b3fa0-c089-4be2-99a6-79c75a148e72/userFiles-b6e5ce16-3f06-4307-9729-29312ef94e15/org.spark-project.spark_unused-1.0.0.jar
23/02/11 07:39:59 INFO spark.SparkContext: Added file file:///home/hadoop/.ivy2/jars/org.lz4_lz4-java-1.4.0.jar at file:///home/hadoop/.ivy2/jars/org.lz4_lz4-java-1.4.0.jar with timestamp 1676101199002
23/02/11 07:39:59 INFO util.Utils: Copying /home/hadoop/.ivy2/jars/org.lz4_lz4-java-1.4.0.jar to /mnt/tmp/spark-3f8b3fa0-c089-4be2-99a6-79c75a148e72/userFiles-b6e5ce16-3f06-4307-9729-29312ef94e15/org.lz4_lz4-java-1.4.0.jar
23/02/11 07:39:59 INFO spark.SparkContext: Added file file:///home/hadoop/.ivy2/jars/org.xerial.snappy_snappy-java-1.1.7.3.jar at file:///home/hadoop/.ivy2/jars/org.xerial.snappy_snappy-java-1.1.7.3.jar with timestamp 1676101199031
23/02/11 07:39:59 INFO util.Utils: Copying /home/hadoop/.ivy2/jars/org.xerial.snappy_snappy-java-1.1.7.3.jar to /mnt/tmp/spark-3f8b3fa0-c089-4be2-99a6-79c75a148e72/userFiles-b6e5ce16-3f06-4307-9729-29312ef94e15/org.xerial.snappy_snappy-java-1.1.7.3.jar
23/02/11 07:39:59 INFO spark.SparkContext: Added file file:///home/hadoop/.ivy2/jars/org.slf4j_slf4j-api-1.7.16.jar at file:///home/hadoop/.ivy2/jars/org.slf4j_slf4j-api-1.7.16.jar with timestamp 1676101199060
```

```
hadoop@ip-172-31-41-58:~
dler@eeda119{/SQL,null,AVAILABLE,@Spark}
23/02/11 07:40:00 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHan
dler@2b5727de{/SQL/json,null,AVAILABLE,@Spark}
23/02/11 07:40:00 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHan
dler@4865ff2a{/SQL/execution,null,AVAILABLE,@Spark}
23/02/11 07:40:00 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHan
dler@7bc6e031{/SQL/execution/json,null,AVAILABLE,@Spark}
23/02/11 07:40:00 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHan
dler@466d0ec4{/static/sql,null,AVAILABLE,@Spark}
23/02/11 07:40:01 INFO state.StateStoreCoordinatorRef: Registered StateStoreCoor
dinator endpoint

-----
Batch: 0
-----

+-----+-----+-----+-----+-----+-----+
|invoice_no|country|timestamp|total_cost|total_items|is_order|is_return|
+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+

-----
Batch: 1
-----

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

```
hadoop@ip-172-31-41-58:~
Batch: 0
-----

+-----+-----+-----+-----+-----+-----+
|invoice_no|country|timestamp|total_cost|total_items|is_order|is_return|
+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+

-----
Batch: 1
-----

+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+
|invoice_no|country|timestamp|total_cost|total_items|is_or
der|is_return|
+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+
|154132553187897|United Kingdom|2023-02-11 07:39:22|5.1000004|6|1
|0|
|154132553187898|United Kingdom|2023-02-11 07:39:31|47.690002|10|1
|0|
|154132553187899|United Kingdom|2023-02-11 07:39:53|10.95|4|1
|0|
|154132553187900|United Kingdom|2023-02-11 07:39:54|74.89|8|1
|0|
```



```
hadoop@ip-172-31-41-58:~
|154132553187905|United Kingdom|2023-02-11 07:40:13|28.68      |21      |1
|0|
|154132553187906|United Kingdom|2023-02-11 07:40:22|238.21     |155     |1
|0|
|154132553187907|United Kingdom|2023-02-11 07:40:25|-39.73   |13      |0
|1|
|154132553187908|United Kingdom|2023-02-11 07:40:28|2.08     |3       |1
|0|
|154132553187909|United Kingdom|2023-02-11 07:40:29|11.11    |27      |1
|0|
|154132553187910|United Kingdom|2023-02-11 07:40:43|37.34    |14      |1
|0|
+-----+-----+-----+-----+-----+-----+
Batch: 2
+-----+-----+-----+-----+-----+-----+
|invoice_no   |country      |timestamp      |total_cost|total_items|is_or
der|is_return|
+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+
```

```
hadoop@ip-172-31-41-58:~
|154132553187935|United Kingdom|2023-02-11 07:41:57|75.59     |17      |1
|0|
|154132553187936|United Kingdom|2023-02-11 07:42:00|24.25     |10      |1
|0|
|154132553187937|United Kingdom|2023-02-11 07:42:25|29.84     |4       |1
|0|
|154132553187938|United Kingdom|2023-02-11 07:42:35|40.06     |21      |1
|0|
|154132553187939|United Kingdom|2023-02-11 07:42:38|34.649998 |7       |1
|0|
|154132553187940|United Kingdom|2023-02-11 07:42:39|0.78      |2       |1
|0|
|154132553187941|United Kingdom|2023-02-11 07:42:39|31.949999 |8       |1
|0|
|154132553187942|United Kingdom|2023-02-11 07:42:40|37.25     |508     |1
|0|
|154132553187943|United Kingdom|2023-02-11 07:42:44|449.92    |176     |1
|0|
|154132553187944|United Kingdom|2023-02-11 07:42:51|18.619999 |4       |1
|0|
+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+
```

```
hadoop@ip-172-31-41-58:~
Batch: 3
-----
+-----+-----+-----+-----+-----+-----+
|invoice_no|country|timestamp|total_cost|total_items|is_order|is_return|
+-----+-----+-----+-----+-----+-----+
|154132553187927|United Kingdom|2023-02-11 07:41:27|45.69|5|1|0|
|154132553187928|United Kingdom|2023-02-11 07:41:28|23.849998|3|1|0|
|154132553187929|United Kingdom|2023-02-11 07:41:30|64.509995|27|1|0|
|154132553187930|United Kingdom|2023-02-11 07:41:55|55.199997|36|1|0|
|154132553187931|United Kingdom|2023-02-11 07:41:59|9.9|6|1|0|
|154132553187932|United Kingdom|2023-02-11 07:41:54|41.149998|19|1|0|
|154132553187933|United Kingdom|2023-02-11 07:41:54|134.28|44|1|0|
|154132553187934|United Kingdom|2023-02-11 07:41:55|1.65|1|1|0|
```

```
hadoop@ip-172-31-41-58:~
Batch: 4
-----
+-----+-----+-----+-----+-----+-----+
|invoice_no|country|timestamp|total_cost|total_items|is_order|is_return|
+-----+-----+-----+-----+-----+-----+
|154132553187945|United Kingdom|2023-02-11 07:42:26|108.16|29|1|0|
|154132553187946|United Kingdom|2023-02-11 07:42:30|24.72|6|1|0|
|154132553187947|United Kingdom|2023-02-11 07:42:34|14.940001|6|1|0|
|154132553187948|United Kingdom|2023-02-11 07:42:36|20.25|3|1|0|
|154132553187949|United Kingdom|2023-02-11 07:42:36|20.8|10|1|0|
|154132553187950|United Kingdom|2023-02-11 07:42:44|137.48|50|1|0|
```

```
hadoop@ip-172-31-41-58:~
|154132553188987|Italy          |2023-02-11 08:36:25|-836.9    |194    |0
|1|
|154132553188988|United Kingdom|2023-02-11 08:36:27|17.7    |6      |1
|0|
|154132553188989|France        |2023-02-11 08:36:03|12.77   |33     |1
|0|
|154132553188990|Portugal     |2023-02-11 08:36:10|47.97   |27     |1
|0|
|154132553188991|Netherlands  |2023-02-11 08:36:23|291.45  |79     |1
|0|
|154132553188992|United Kingdom|2023-02-11 08:36:26|36.29   |15     |1
|0|
|154132553188993|United Kingdom|2023-02-11 08:36:30|16.99   |3      |1
|0|
|154132553188994|United Kingdom|2023-02-11 08:36:31|4.13    |1      |1
|0|
|154132553188995|United Kingdom|2023-02-11 08:36:32|8.29    |1      |1
|0|
|154132553188996|United Kingdom|2023-02-11 08:36:48|298.1   |54     |1
|0|
+-----+-----+-----+-----+-----+
+-----+
```

```
hadoop@ip-172-31-41-58:~
|154132553188996|United Kingdom|2023-02-11 08:36:48|298.1   |54     |1
|0|
+-----+-----+-----+-----+-----+
+-----+

Batch: 5
+-----+
+-----+
|invoice_no|country      |timestamp          |total_cost|total_items|is_order|is_return|
+-----+-----+-----+-----+-----+-----+
+-----+
|154132553188997|United Kingdom|2023-02-11 08:36:29|168.36   |48      |1
|0|
|154132553188998|United Kingdom|2023-02-11 08:36:44|2.1      |1       |1
|0|
|154132553188999|United Kingdom|2023-02-11 08:36:47|251.6    |39      |1
|0|
|154132553189000|United Kingdom|2023-02-11 08:37:00|4.94     |7       |1
|0|
|154132553189001|United Kingdom|2023-02-11 08:37:03|53.1     |18      |1
|0|
+-----+-----+-----+-----+-----+
+-----+
```

hadoop@ip-172-31-41-58:~

invoice_no	country	timestamp	total_cost	total_items	is_order	is_return
154132553188997	United Kingdom	2023-02-11 08:36:29	168.36	48	1	1
154132553188998	United Kingdom	2023-02-11 08:36:44	2.1	1	1	1
154132553188999	United Kingdom	2023-02-11 08:36:47	251.6	39	1	1
154132553189000	United Kingdom	2023-02-11 08:37:00	4.94	7	1	1
154132553189001	United Kingdom	2023-02-11 08:37:03	53.1	18	1	1
154132553189002	United Kingdom	2023-02-11 08:37:15	6.52	4	1	1
154132553189003	United Kingdom	2023-02-11 08:37:12	98.76	46	1	1
154132553189004	France	2023-02-11 08:37:22	28.7	9	1	1
154132553189005	Unspecified	2023-02-11 08:37:22	2.89	1	1	1
154132553189006	United Kingdom	2023-02-11 08:37:23	8.26	2	1	1
154132553189007	United Kingdom	2023-02-11 08:37:28	157.95	39	1	1

hadoop@ip-172-31-41-58:~

Batch: 6

invoice_no	country	timestamp	total_cost	total_items	is_order	is_return
154132553189012	United Kingdom	2023-02-11 08:37:38	121.11	69	1	1
154132553189013	United Kingdom	2023-02-11 08:38:05	19.26	3	1	1
154132553189014	United Kingdom	2023-02-11 08:38:14	18.279999	5	1	1
154132553189015	United Kingdom	2023-02-11 08:38:16	19.38	13	1	1
154132553189016	United Kingdom	2023-02-11 08:38:18	88.38	50	1	1
154132553189017	United Kingdom	2023-02-11 08:38:18	-34.73	21	0	0
154132553189018	United Kingdom	2023-02-11 08:38:02	7.3	4	1	1
154132553189019	United Kingdom	2023-02-11 08:38:07	0.55	1	1	1

```
hadoop@ip-172-31-41-58:~
Batch: 7
-----
+-----+-----+-----+-----+-----+-----+
|invoice_no|country|timestamp|total_cost|total_items|is_order|is_return|
+-----+-----+-----+-----+-----+-----+
|154132553189026|United Kingdom|2023-02-11 08:38:27|93.89|37|1|1|
|0|
|154132553189027|United Kingdom|2023-02-11 08:38:30|19.35|7|1|1|
|0|
|154132553189028|United Kingdom|2023-02-11 08:38:40|-44.55|9|0|0|
|1|
|154132553189029|United Kingdom|2023-02-11 08:38:50|-30.73|6|0|0|
|1|
|154132553189030|United Kingdom|2023-02-11 08:38:54|62.850002|17|1|1|
|0|
|154132553189031|United Kingdom|2023-02-11 08:38:55|178.45|67|1|1|
|0|
|154132553189032|United Kingdom|2023-02-11 08:38:57|0.85|1|1|1|
|0|
|154132553189033|United Kingdom|2023-02-11 08:38:59|9.96|12|1|1|
```

```
hadoop@ip-172-31-41-58:~
Batch: 8
-----
+-----+-----+-----+-----+-----+-----+
|invoice_no|country|timestamp|total_cost|total_items|is_order|is_return|
+-----+-----+-----+-----+-----+-----+
|154132553189046|United Kingdom|2023-02-11 08:39:39|42.55|25|1|1|
|0|
|154132553189047|United Kingdom|2023-02-11 08:39:39|15.34|8|1|1|
|0|
|154132553189048|United Kingdom|2023-02-11 08:39:43|44.33|37|1|1|
|0|
|154132553189049|United Kingdom|2023-02-11 08:39:47|178.83|79|1|1|
|0|
|154132553189050|United Kingdom|2023-02-11 08:39:50|-141.98|38|0|0|
|1|
|154132553189051|United Kingdom|2023-02-11 08:39:50|35.75|11|1|1|
|0|
|154132553189052|United Kingdom|2023-02-11 08:39:53|77.76|22|1|1|
|0|
|154132553189053|United Kingdom|2023-02-11 08:39:59|12.599999|6|1|1|
```

```
hadoop@ip-172-31-41-58:~
Batch: 9
-----
+-----+-----+-----+-----+-----+-----+
|invoice_no|country|timestamp|total_cost|total_items|is_order|is_return|
+-----+-----+-----+-----+-----+-----+
|154132553189069|Germany|2023-02-11 08:40:37|15.25|6|1|1|
|0|
|154132553189070|France|2023-02-11 08:40:43|116.56|56|1|1|
|0|
|154132553189071|United Kingdom|2023-02-11 08:40:49|91.92|17|1|1|
|0|
|154132553189072|United Kingdom|2023-02-11 08:40:53|89.38|19|1|1|
|0|
|154132553189073|United Kingdom|2023-02-11 08:40:58|112.549995|2|1|1|
|0|
|154132553189074|United Kingdom|2023-02-11 08:40:59|46.4|12|1|1|
|0|
|154132553189075|United Kingdom|2023-02-11 08:40:59|97.45|46|1|1|
|0|
|154132553189076|United Kingdom|2023-02-11 08:41:12|-16.65|13|0|0|
|1|
|1|
```

```
hadoop@ip-172-31-41-58:~
Batch: 10
-----
+-----+-----+-----+-----+-----+-----+
|invoice_no|country|timestamp|total_cost|total_items|is_order|is_return|
+-----+-----+-----+-----+-----+-----+
|154132553189081|United Kingdom|2023-02-11 08:41:25|343.0|77|1|1|
|0|
|154132553189082|United Kingdom|2023-02-11 08:41:33|40.800003|48|1|1|
|0|
|154132553189083|United Kingdom|2023-02-11 08:41:42|60.089996|31|1|1|
|0|
|154132553189084|United Kingdom|2023-02-11 08:42:19|36.739998|14|1|1|
|0|
|154132553189085|United Kingdom|2023-02-11 08:41:58|25.25|7|1|1|
|0|
|154132553189086|United Kingdom|2023-02-11 08:41:59|20.650002|5|1|1|
|0|
|154132553189087|United Kingdom|2023-02-11 08:42:02|42.42|28|1|1|
|0|
|154132553189088|United Kingdom|2023-02-11 08:42:04|17.55|13|1|1|
|0|
|0|
```

```
hadoop@ip-172-31-41-58:~
Batch: 11
-----
+-----+-----+-----+-----+-----+-----+
|invoice_no|country|timestamp|total_cost|total_items|is_order|is_return|
+-----+-----+-----+-----+-----+-----+
|154132553189099|United Kingdom|2023-02-11 08:42:22|20.82|14|1|1|
|0|
|154132553189100|United Kingdom|2023-02-11 08:42:23|34.77|53|1|1|
|0|
|154132553189101|United Kingdom|2023-02-11 08:42:30|19.17|15|1|1|
|0|
|154132553189102|United Kingdom|2023-02-11 08:42:38|15.57|51|1|1|
|0|
|154132553189103|United Kingdom|2023-02-11 08:42:43|165.89|49|1|1|
|0|
|154132553189104|United Kingdom|2023-02-11 08:42:59|87.060005|66|1|1|
|0|
|154132553189105|United Kingdom|2023-02-11 08:43:06|3.84|2|1|1|
|0|
|154132553189106|United Kingdom|2023-02-11 08:43:08|19.96|1|1|1|
|0|
```

```
hadoop@ip-172-31-41-58:~
Batch: 12
-----
+-----+-----+-----+-----+-----+-----+
|invoice_no|country|timestamp|total_cost|total_items|is_order|is_return|
+-----+-----+-----+-----+-----+-----+
|154132553189118|United Kingdom|2023-02-11 08:43:39|-15.9|2|0|0|
|1|
|154132553189119|United Kingdom|2023-02-11 08:43:44|52.25|21|1|1|
|0|
|154132553189120|United Kingdom|2023-02-11 08:43:49|0.83|1|1|1|
|0|
|154132553189121|United Kingdom|2023-02-11 08:43:50|40.38|42|1|1|
|0|
|154132553189122|United Kingdom|2023-02-11 08:44:02|102.25|17|1|1|
|0|
|154132553189123|United Kingdom|2023-02-11 08:44:08|50.1|19|1|1|
|0|
|154132553189124|United Kingdom|2023-02-11 08:44:11|-25.42|9|0|0|
|1|
|154132553189125|United Kingdom|2023-02-11 08:44:19|152.72|112|1|1|
```

```
hadoop@ip-172-31-41-58:~
Batch: 13
-----
+-----+-----+-----+-----+-----+-----+
|invoice_no|country|timestamp|total_cost|total_items|is_order|is_return|
+-----+-----+-----+-----+-----+-----+
|154132553189132|United Kingdom|2023-02-11 08:44:21|144.22|73|1|1|
|0|
|154132553189133|United Kingdom|2023-02-11 08:44:30|45.03|6|1|1|
|0|
|154132553189134|United Kingdom|2023-02-11 08:44:31|100.64|110|1|1|
|0|
|154132553189135|United Kingdom|2023-02-11 08:44:33|36.17|12|1|1|
|0|
|154132553189136|United Kingdom|2023-02-11 08:44:39|24.75|61|1|1|
|0|
|154132553189137|United Kingdom|2023-02-11 08:44:45|20.599998|6|1|1|
|0|
|154132553189138|United Kingdom|2023-02-11 08:44:46|7.5|6|1|1|
|0|
|154132553189139|United Kingdom|2023-02-11 08:44:50|408.07|172|1|1|
```

```
hadoop@ip-172-31-41-58:~
Batch: 14
-----
+-----+-----+-----+-----+-----+-----+
|invoice_no|country|timestamp|total_cost|total_items|is_order|is_return|
+-----+-----+-----+-----+-----+-----+
|154132553189156|United Kingdom|2023-02-11 08:45:20|44.76|16|1|1|
|0|
|154132553189157|United Kingdom|2023-02-11 08:45:20|12.48|6|1|1|
|0|
|154132553189158|United Kingdom|2023-02-11 08:45:32|-38.19|28|0|0|
|1|
|154132553189159|United Kingdom|2023-02-11 08:45:41|10.95|7|1|1|
|0|
|154132553189160|United Kingdom|2023-02-11 08:45:51|152.04|72|1|1|
|0|
|154132553189161|United Kingdom|2023-02-11 08:46:00|3.32|1|1|1|
|0|
|154132553189162|United Kingdom|2023-02-11 08:46:03|49.56|12|1|1|
|0|
|154132553189163|United Kingdom|2023-02-11 08:46:04|-44.06|7|0|0|
```



```
hadoop@ip-172-31-41-58:~
Batch: 15
-----
+-----+-----+-----+-----+-----+-----+
|invoice_no|country|timestamp|total_cost|total_items|is_order|is_return|
+-----+-----+-----+-----+-----+-----+
|154132553189172|United Kingdom|2023-02-11 08:46:18|27.85|49|1|0|
|154132553189173|United Kingdom|2023-02-11 08:46:19|30.14|16|1|0|
|154132553189174|United Kingdom|2023-02-11 08:46:21|22.2|16|1|0|
|154132553189175|United Kingdom|2023-02-11 08:46:25|74.1|26|1|0|
|154132553189176|United Kingdom|2023-02-11 08:46:26|30.0|24|1|0|
|154132553189177|United Kingdom|2023-02-11 08:46:29|135.3|77|1|0|
|154132553189178|United Kingdom|2023-02-11 08:46:33|28.72|14|1|0|
|154132553189179|United Kingdom|2023-02-11 08:46:48|31.159998|8|1|0|
```

```
hadoop@ip-172-31-41-58:~
Batch: 16
-----
+-----+-----+-----+-----+-----+-----+
|invoice_no|country|timestamp|total_cost|total_items|is_order|is_return|
+-----+-----+-----+-----+-----+-----+
|154132553189194|United Kingdom|2023-02-11 08:47:23|55.2|22|1|0|
|154132553189195|United Kingdom|2023-02-11 08:47:38|51.739998|20|1|0|
|154132553189196|United Kingdom|2023-02-11 08:47:53|25.25|13|1|0|
|154132553189197|United Kingdom|2023-02-11 08:47:56|80.33|27|1|0|
|154132553189198|United Kingdom|2023-02-11 08:48:00|140.75|111|1|0|
|154132553189199|United Kingdom|2023-02-11 08:48:02|76.049995|16|1|0|
|154132553189200|United Kingdom|2023-02-11 08:48:04|14.78|3|1|0|
|154132553189201|United Kingdom|2023-02-11 08:48:07|16.6|4|1|0|
```

```
hadoop@ip-172-31-41-58:~
Batch: 17
-----
+-----+-----+-----+-----+-----+-----+
|invoice_no|country|timestamp|total_cost|total_items|is_order|is_return|
+-----+-----+-----+-----+-----+-----+
|154132553189214|United Kingdom|2023-02-11 08:48:16|35.0|32|1|1|
|0|
|154132553189215|United Kingdom|2023-02-11 08:48:21|85.51|24|1|1|
|0|
|154132553189216|United Kingdom|2023-02-11 08:48:29|49.53|28|1|1|
|0|
|154132553189217|United Kingdom|2023-02-11 08:48:30|24.75|3|1|1|
|0|
|154132553189218|United Kingdom|2023-02-11 08:48:35|78.479996|41|1|1|
|0|
|154132553189219|United Kingdom|2023-02-11 08:48:38|7.7999997|12|1|1|
|0|
|154132553189220|France|2023-02-11 08:48:47|74.98|26|1|1|
|0|
|154132553189221|United Kingdom|2023-02-11 08:48:50|52.05|35|1|1|
|0|
```

```
hadoop@ip-172-31-41-58:~
Batch: 18
-----
+-----+-----+-----+-----+-----+-----+
|invoice_no|country|timestamp|total_cost|total_items|is_order|is_return|
+-----+-----+-----+-----+-----+-----+
|154132553189232|United Kingdom|2023-02-11 08:49:19|21.18|26|1|1|
|0|
|154132553189233|United Kingdom|2023-02-11 08:49:30|19.98|4|1|1|
|0|
|154132553189234|United Kingdom|2023-02-11 08:49:31|30.88|10|1|1|
|0|
|154132553189235|United Kingdom|2023-02-11 08:49:32|1.25|1|1|1|
|0|
|154132553189236|United Kingdom|2023-02-11 08:49:36|6.2999997|3|1|1|
|0|
|154132553189237|United Kingdom|2023-02-11 08:49:39|42.66|13|1|1|
|0|
|154132553189238|United Kingdom|2023-02-11 08:49:43|3.36|1|1|1|
|0|
|154132553189239|United Kingdom|2023-02-11 08:49:50|28.78|22|1|1|
|0|
```

```
hadoop@ip-172-31-41-58:~
Batch: 19
-----
+-----+-----+-----+-----+-----+-----+
+-----+
|invoice_no|country|timestamp|total_cost|total_items|is_order|is_return|
+-----+-----+-----+-----+-----+-----+
+-----+
|154132553189256|United Kingdom|2023-02-11 08:50:17|16.05|3|1|0|
|154132553189257|United Kingdom|2023-02-11 08:50:34|18.38|8|1|0|
|154132553189258|United Kingdom|2023-02-11 08:50:34|5.0099998|2|1|0|
|154132553189259|United Kingdom|2023-02-11 08:50:53|50.25|31|1|0|
|154132553189260|United Kingdom|2023-02-11 08:50:48|24.900002|6|1|0|
|154132553189261|United Kingdom|2023-02-11 08:50:55|24.03|17|1|0|
|154132553189262|United Kingdom|2023-02-11 08:50:56|37.11|27|1|0|
|154132553189263|United Kingdom|2023-02-11 08:51:06|6.2000003|6|1|0|
```

```
hadoop@ip-172-31-41-58:~
Batch: 20
-----
+-----+-----+-----+-----+-----+-----+
+-----+
|invoice_no|country|timestamp|total_cost|total_items|is_order|is_return|
+-----+-----+-----+-----+-----+-----+
+-----+
|154132553189269|United Kingdom|2023-02-11 08:51:26|829.08|252|1|0|
|154132553189270|United Kingdom|2023-02-11 08:51:26|374.13998|134|1|0|
|154132553189271|United Kingdom|2023-02-11 08:51:27|42.64|16|1|0|
|154132553189272|United Kingdom|2023-02-11 08:51:29|84.53|74|1|0|
|154132553189273|United Kingdom|2023-02-11 08:51:34|33.38|6|1|0|
|154132553189274|United Kingdom|2023-02-11 08:51:41|55.59|15|1|0|
|154132553189275|United Kingdom|2023-02-11 08:51:41|5.65|3|1|0|
|154132553189276|United Kingdom|2023-02-11 08:51:52|41.42|9|1|0|
```

```
hadoop@ip-172-31-41-58:~
|154132553189278|United Kingdom|2023-02-11 08:52:03|564.68      |200      |1
|0              |
|154132553189279|United Kingdom|2023-02-11 08:51:47|50.74      |36        |1
|0              |
|154132553189280|United Kingdom|2023-02-11 08:52:00|30.36      |8          |1
|0              |
|154132553189281|United Kingdom|2023-02-11 08:52:15|28.6       |14         |1
|0              |
|154132553189282|United Kingdom|2023-02-11 08:52:17|-2.1       |1          |0
|1              |
|154132553189283|United Kingdom|2023-02-11 08:52:22|23.82      |13         |1
|0              |
|154132553189284|United Kingdom|2023-02-11 08:52:26|-12.75     |1          |0
|1              |
|154132553189285|United Kingdom|2023-02-11 08:52:43|59.34      |18         |1
|0              |
+-----+-----+-----+-----+-----+-----+
---+-----+

^CTraceback (most recent call last):
  File "/home/hadoop/spark-streaming.py", line 141, in <module>
    ByTime_country.awaitTermination()
  File "/usr/lib/spark/python/lib/pyspark.zip/pyspark/sql/streaming.py", line 10
3, in awaitTermination
```

```
[hadoop@ip-172-31-88-120 ~]$ hadoop fs -ls
Found 3 items
drwxr-xr-x  - hadoop hadoop      0 2023-02-07 10:54 .sparkStaging
drwxr-xr-x  - hadoop hadoop      0 2023-02-07 10:32 country_kpi
drwxr-xr-x  - hadoop hadoop      0 2023-02-07 10:32 time_kpi
[hadoop@ip-172-31-88-120 ~]$
```

```
hadoop@ip-172-31-88-120~$ hadoop fs -ls country_kpi
Found 53 items
drwxr-xr-x - hadoop hadoop 0 2023-02-07 10:32 country_kpi/_spark_metadata
drwxr-xr-x - hadoop hadoop 0 2023-02-07 10:12 country_kpi/cp
-rw-r--r-- 1 hadoop hadoop 0 2023-02-07 10:23 country_kpi/part-00000-040fbeb1-5faa-4d13-be43-2af7417802e-c
000.json
-rw-r--r-- 1 hadoop hadoop 0 2023-02-07 10:12 country_kpi/part-00000-04350840-c873-4c25-841b-cfbe03ba76cb-c
000.json
-rw-r--r-- 1 hadoop hadoop 0 2023-02-07 10:27 country_kpi/part-00000-13da4873-dc66-4067-866b-dbd9f2beffbb-c
000.json
-rw-r--r-- 1 hadoop hadoop 0 2023-02-07 10:29 country_kpi/part-00000-1e41d30f-cc95-4ac7-9ded-86e7ce1bb38e-c
000.json
-rw-r--r-- 1 hadoop hadoop 0 2023-02-07 10:12 country_kpi/part-00000-1fef0062-8f5f-4b0b-8f3b-fc6760a6e9c9-c
000.json
-rw-r--r-- 1 hadoop hadoop 0 2023-02-07 10:28 country_kpi/part-00000-29a27fbc-c0f1-498e-a773-5b496c03a6bc-c
000.json
-rw-r--r-- 1 hadoop hadoop 0 2023-02-07 10:31 country_kpi/part-00000-33efbe28-f054-4859-bcde-d85ed74cb395-c
000.json
-rw-r--r-- 1 hadoop hadoop 0 2023-02-07 10:20 country_kpi/part-00000-44ae2c66-2ebb-4b17-b233-46a91a5b3f1e-c
000.json
-rw-r--r-- 1 hadoop hadoop 0 2023-02-07 10:17 country_kpi/part-00000-5a4687be-b3f0-44c0-860d-2f97256d3cae-c
000.json
-rw-r--r-- 1 hadoop hadoop 0 2023-02-07 10:16 country_kpi/part-00000-79840255-73c3-4328-9fa4-1f2bd0408fa6-c
000.json
-rw-r--r-- 1 hadoop hadoop 0 2023-02-07 10:22 country_kpi/part-00000-85012f25-960b-4990-89cb-68df85927c11-c
000.json
-rw-r--r-- 1 hadoop hadoop 0 2023-02-07 10:15 country_kpi/part-00000-9b5e05fd-a061-4a1c-abdc-882fb75dc9f7-c
000.json
-rw-r--r-- 1 hadoop hadoop 0 2023-02-07 10:24 country_kpi/part-00000-ab1b18cd-c5f8-40d5-989f-08a0ac1ccb64-c
000.json
-rw-r--r-- 1 hadoop hadoop 0 2023-02-07 10:14 country_kpi/part-00000-acd9a0f0-3471-4278-8cf6-2767a3171279-c
000.json
-rw-r--r-- 1 hadoop hadoop 0 2023-02-07 10:30 country_kpi/part-00000-b7a97bd6-b9ad-4e1f-b04d-b9fa84b8ec4d-c
000.json
```

```
hadoop@ip-172-31-88-120~$ hadoop fs -ls country_kpi
-rw-r--r-- 1 hadoop hadoop 169 2023-02-07 10:20 country_kpi/part-00001-90f7125e-4b1b-4c5f-b573-f78b11d82eda-c
000.json
-rw-r--r-- 1 hadoop hadoop 184 2023-02-07 10:23 country_kpi/part-00006-5bbd2e37-9cf8-4310-9a32-e0ab581c0ab4-c
000.json
-rw-r--r-- 1 hadoop hadoop 168 2023-02-07 10:30 country_kpi/part-00009-b3db694e-4026-4378-b43a-27ee557b3615-c
000.json
-rw-r--r-- 1 hadoop hadoop 161 2023-02-07 10:28 country_kpi/part-00020-c0c3aa14-4d87-448c-8975-6fd74394dfae-c
000.json
-rw-r--r-- 1 hadoop hadoop 169 2023-02-07 10:27 country_kpi/part-00027-ded14fb8-fbef-4407-a963-88322f656233-c
000.json
-rw-r--r-- 1 hadoop hadoop 158 2023-02-07 10:18 country_kpi/part-00040-9dcce1df-ce90-4de0-8ac4-1d3efcf38449-c
000.json
-rw-r--r-- 1 hadoop hadoop 158 2023-02-07 10:15 country_kpi/part-00041-6861db82-6dca-43f5-8609-a21e7277e43e-c
000.json
-rw-r--r-- 1 hadoop hadoop 168 2023-02-07 10:18 country_kpi/part-00044-e22360f2-721d-402e-b4c3-3b464ed163f3-c
000.json
-rw-r--r-- 1 hadoop hadoop 169 2023-02-07 10:24 country_kpi/part-00053-b6a2db24-e850-47b8-8c2a-50fbddfd4fac-c
000.json
-rw-r--r-- 1 hadoop hadoop 162 2023-02-07 10:16 country_kpi/part-00085-10a972de-cdc5-49f9-a5bf-42b9c6317974-c
000.json
-rw-r--r-- 1 hadoop hadoop 162 2023-02-07 10:18 country_kpi/part-00093-ea95b3cb-52fb-4f18-bd3c-52860702c8a5-c
000.json
-rw-r--r-- 1 hadoop hadoop 185 2023-02-07 10:15 country_kpi/part-00100-51310825-cebe-4bfb-a184-2f2cc39c2aed-c
000.json
-rw-r--r-- 1 hadoop hadoop 163 2023-02-07 10:21 country_kpi/part-00102-fe6d6fa1-e0ed-4fbe-81cb-4f9416953886-c
000.json
-rw-r--r-- 1 hadoop hadoop 169 2023-02-07 10:25 country_kpi/part-00109-35ca7351-7f5e-4271-84f0-fb17d8a647f7-c
000.json
-rw-r--r-- 1 hadoop hadoop 148 2023-02-07 10:21 country_kpi/part-00114-5d561de3-97e9-41b6-83e1-7535e58edecf-c
000.json
-rw-r--r-- 1 hadoop hadoop 170 2023-02-07 10:16 country_kpi/part-00115-62e1e6f7-e4f0-4578-b2eb-3125741ef87a-c
000.json
-rw-r--r-- 1 hadoop hadoop 169 2023-02-07 10:18 country_kpi/part-00126-7c489bcc-7fd6-4f5e-95e3-a2e3efe506d8-c
000.json
```



```
hadoop@ip-172-31-88-120~$
000.json
-rw-r--r-- 1 hadoop hadoop 170 2023-02-07 10:16 country_kpi/part-00115-62e1e6f7-e4f0-4578-b2eb-3125741ef87a-c
000.json
-rw-r--r-- 1 hadoop hadoop 169 2023-02-07 10:18 country_kpi/part-00126-7c489bcc-7fd6-4f5e-95e3-a2e3efe506d8-c
000.json
-rw-r--r-- 1 hadoop hadoop 161 2023-02-07 10:23 country_kpi/part-00127-3ae7ed0e-7aed-4837-84fb-9df75ef3ad82-c
000.json
-rw-r--r-- 1 hadoop hadoop 169 2023-02-07 10:18 country_kpi/part-00133-db0ad497-ccb9-402b-b662-c63b28ce9042-c
000.json
-rw-r--r-- 1 hadoop hadoop 170 2023-02-07 10:29 country_kpi/part-00138-4db99e81-50a9-4114-b087-f117784a0450-c
000.json
-rw-r--r-- 1 hadoop hadoop 160 2023-02-07 10:18 country_kpi/part-00146-41aebb58-d109-4195-b610-e006f24bd167-c
000.json
-rw-r--r-- 1 hadoop hadoop 185 2023-02-07 10:26 country_kpi/part-00151-ab53f162-55ad-40d9-bff1-e952d36dd7b3-c
000.json
-rw-r--r-- 1 hadoop hadoop 185 2023-02-07 10:21 country_kpi/part-00155-238d365a-5a5e-4b04-850a-cb55b038918e-c
000.json
-rw-r--r-- 1 hadoop hadoop 169 2023-02-07 10:19 country_kpi/part-00157-5ea5f5de-e9ea-4e3a-8d3f-ce4b8c85f565-c
000.json
-rw-r--r-- 1 hadoop hadoop 169 2023-02-07 10:32 country_kpi/part-00165-eca4a3ff-1f4e-4be9-841e-c2d70f4f1fe-c
000.json
-rw-r--r-- 1 hadoop hadoop 166 2023-02-07 10:28 country_kpi/part-00166-c132be76-874c-48d6-93ff-58ab5abc28f1-c
000.json
-rw-r--r-- 1 hadoop hadoop 160 2023-02-07 10:20 country_kpi/part-00174-4386b5c9-049e-4709-a626-00ee875d955b-c
000.json
-rw-r--r-- 1 hadoop hadoop 185 2023-02-07 10:28 country_kpi/part-00179-20906c0d-c781-4ed1-ad1c-2f8f02c81c70-c
000.json
-rw-r--r-- 1 hadoop hadoop 185 2023-02-07 10:22 country_kpi/part-00199-567904be-f6ea-410f-b54e-fe75efca0a66-c
000.json
[hadoop@ip-172-31-88-120 ~]$ hadoop fs -cat country_kpi/part-00199-567904be-f6ea-410f-b54e-fe75efca0a66-c000.json
{"start":"2023-02-07T10:18:00.000Z","end":"2023-02-07T10:19:00.000Z","country":"United Kingdom","OPM":11,"total_volume
of_sales":258.1399978399277,"rate_of_return":0.09090909090909091}
[hadoop@ip-172-31-88-120 ~]$
```

```
hadoop@ip-172-31-88-120~$
[hadoop@ip-172-31-88-120 ~]$ hadoop fs -ls time_kpi
Found 41 items
drwxr-xr-x - hadoop hadoop 0 2023-02-07 10:32 time_kpi/ spark_metadata
drwxr-xr-x - hadoop hadoop 0 2023-02-07 10:12 time_kpi/cp
-rw-r--r-- 1 hadoop hadoop 0 2023-02-07 10:28 time_kpi/part-00000-2214fd79-d3a9-4479-9f77-e41e2970c939-c000
.json
-rw-r--r-- 1 hadoop hadoop 0 2023-02-07 10:23 time_kpi/part-00000-340cbb83-4bdc-48e1-b7e5-2378123a88ff-c000
.json
-rw-r--r-- 1 hadoop hadoop 0 2023-02-07 10:14 time_kpi/part-00000-3dbf7151-7ff2-47af-a88c-4bb776b8717c-c000
.json
-rw-r--r-- 1 hadoop hadoop 0 2023-02-07 10:24 time_kpi/part-00000-40577bb3-241c-4e2e-8128-a45ba68d0596-c000
.json
-rw-r--r-- 1 hadoop hadoop 0 2023-02-07 10:13 time_kpi/part-00000-596a1029-fc20-491b-a7da-9dda17ab06e0-c000
.json
-rw-r--r-- 1 hadoop hadoop 0 2023-02-07 10:18 time_kpi/part-00000-59a3884e-f591-41fd-b476-a8fe3318f5ae-c000
.json
-rw-r--r-- 1 hadoop hadoop 0 2023-02-07 10:27 time_kpi/part-00000-5a46a320-482d-4649-9499-e54ed737e503-c000
.json
-rw-r--r-- 1 hadoop hadoop 0 2023-02-07 10:16 time_kpi/part-00000-5e897bf2-392e-48fd-9a93-59835c1a1677-c000
.json
-rw-r--r-- 1 hadoop hadoop 0 2023-02-07 10:19 time_kpi/part-00000-732628c8-bd1b-428d-89a7-675eb2256fc8-c000
.json
-rw-r--r-- 1 hadoop hadoop 0 2023-02-07 10:32 time_kpi/part-00000-78c2f0ef-1ebc-4f15-a444-7dbc12cb38fe-c000
.json
-rw-r--r-- 1 hadoop hadoop 0 2023-02-07 10:20 time_kpi/part-00000-8033822c-66f9-4488-a60e-37b20d2bcd05-c000
.json
-rw-r--r-- 1 hadoop hadoop 0 2023-02-07 10:26 time_kpi/part-00000-8674c0db-f49e-4425-aae1-030abd487108-c000
.json
-rw-r--r-- 1 hadoop hadoop 0 2023-02-07 10:21 time_kpi/part-00000-a50e69b6-ac33-49db-9bd4-7cb658339c08-c000
.json
-rw-r--r-- 1 hadoop hadoop 0 2023-02-07 10:12 time_kpi/part-00000-ad5a49b4-1f39-4954-aa6e-129386080f71-c000
.json
-rw-r--r-- 1 hadoop hadoop 0 2023-02-07 10:31 time_kpi/part-00000-b3151f22-16c1-4a25-aac5-c0b628d387e3-c000
```

```
hadoop@ip-172-31-88-120:~$ cat time_kpi/part-00042-8b653bc9-0d1c-43f2-ab29-0ca01025ed67-c000.json
-rw-r--r-- 1 hadoop hadoop 179 2023-02-07 10:30 time_kpi/part-00042-8b653bc9-0d1c-43f2-ab29-0ca01025ed67-c000
hadoop@ip-172-31-88-120:~$ cat time_kpi/part-00085-6fec8349-7a90-488e-b9e2-f45775032aa8-c000.json
-rw-r--r-- 1 hadoop hadoop 195 2023-02-07 10:18 time_kpi/part-00085-6fec8349-7a90-488e-b9e2-f45775032aa8-c000
hadoop@ip-172-31-88-120:~$ cat time_kpi/part-00107-29aa824d-eba1-49ee-96b7-2c28e6b45794-c000.json
-rw-r--r-- 1 hadoop hadoop 178 2023-02-07 10:24 time_kpi/part-00107-29aa824d-eba1-49ee-96b7-2c28e6b45794-c000
hadoop@ip-172-31-88-120:~$ cat time_kpi/part-00110-291f7993-3c36-4930-aa44-ed92284978d5-c000.json
-rw-r--r-- 1 hadoop hadoop 179 2023-02-07 10:19 time_kpi/part-00110-291f7993-3c36-4930-aa44-ed92284978d5-c000
hadoop@ip-172-31-88-120:~$ cat time_kpi/part-00114-20ed2e7c-d96c-4666-ae44-573a7923c805-c000.json
-rw-r--r-- 1 hadoop hadoop 180 2023-02-07 10:18 time_kpi/part-00114-20ed2e7c-d96c-4666-ae44-573a7923c805-c000
hadoop@ip-172-31-88-120:~$ cat time_kpi/part-00135-76437f8b-6014-46d2-ab60-76f89adc3cb5-c000.json
-rw-r--r-- 1 hadoop hadoop 178 2023-02-07 10:29 time_kpi/part-00135-76437f8b-6014-46d2-ab60-76f89adc3cb5-c000
hadoop@ip-172-31-88-120:~$ cat time_kpi/part-00154-27566126-3f99-4e6f-ac2a-800b6c5c0cf6-c000.json
-rw-r--r-- 1 hadoop hadoop 179 2023-02-07 10:27 time_kpi/part-00154-27566126-3f99-4e6f-ac2a-800b6c5c0cf6-c000
hadoop@ip-172-31-88-120:~$ cat time_kpi/part-00177-49997d75-5259-4ee7-bfd7-01fb698d145d-c000.json
-rw-r--r-- 1 hadoop hadoop 181 2023-02-07 10:23 time_kpi/part-00177-49997d75-5259-4ee7-bfd7-01fb698d145d-c000
hadoop@ip-172-31-88-120:~$ cat time_kpi/part-00177-5a254433-3bbf-4913-be79-5bc2fc5389db-c000.json
-rw-r--r-- 1 hadoop hadoop 179 2023-02-07 10:32 time_kpi/part-00177-5a254433-3bbf-4913-be79-5bc2fc5389db-c000
hadoop@ip-172-31-88-120:~$ cat time_kpi/part-00186-bf89d10f-d41-4325-ba06-1e2f6730f36f-c000.json
-rw-r--r-- 1 hadoop hadoop 196 2023-02-07 10:21 time_kpi/part-00186-bf89d10f-d41-4325-ba06-1e2f6730f36f-c000
hadoop@ip-172-31-88-120:~$ cat time_kpi/part-00186-ff2b23b9-07ff-4338-936a-225532e971bc-c000.json
-rw-r--r-- 1 hadoop hadoop 194 2023-02-07 10:28 time_kpi/part-00186-ff2b23b9-07ff-4338-936a-225532e971bc-c000
hadoop@ip-172-31-88-120:~$ cat time_kpi/part-00193-80db59a3-6a22-465b-8cb8-55df6ef229e7-c000.json
-rw-r--r-- 1 hadoop hadoop 195 2023-02-07 10:26 time_kpi/part-00193-80db59a3-6a22-465b-8cb8-55df6ef229e7-c000
hadoop@ip-172-31-88-120:~$ cat time_kpi/part-00195-3d6f7ca8-9d53-4e09-8e1b-eec2866aa2eb-c000.json
-rw-r--r-- 1 hadoop hadoop 195 2023-02-07 10:22 time_kpi/part-00195-3d6f7ca8-9d53-4e09-8e1b-eec2866aa2eb-c000
hadoop@ip-172-31-88-120:~$ cat time_kpi/part-00196-57c5eb7f-8e98-40d6-bb77-9803ec7e888d-c000.json
-rw-r--r-- 1 hadoop hadoop 179 2023-02-07 10:16 time_kpi/part-00196-57c5eb7f-8e98-40d6-bb77-9803ec7e888d-c000
hadoop@ip-172-31-88-120:~$ hadoop fs -cat time_kpi/part-00195-3d6f7ca8-9d53-4e09-8e1b-eec2866aa2eb-c000.json
{"start":"2023-02-07T10:18:00.000Z","end":"2023-02-07T10:19:00.000Z","total_volume_of_sales":258.1399978399277,"average_transaction_size":23.467272530902516,"rate_of_return":0.09090909090909091}
hadoop@ip-172-31-88-120:~$
```