

Code Logic - Retail Data Analysis

Step 1. Create folder **Retail_Project** and paste jar file - **spark-sql-kafka-0-10_2.11-2.3.0.jar** in folder on terminal ec2-user.

Step 2. Create python file in the same folder -> vi **spark-streaming.py** . In this file, written the logic to run the file using spark. Check below screenshot for reference.

```
[ec2-user@ip-10-0-0-135 ~]$ cd Retail_Project
[ec2-user@ip-10-0-0-135 Retail_Project]$ ls
Console-output  op2                                spark-streaming.py
op1              spark-sql-kafka-0-10_2.11-2.3.0.jar  tmp
[ec2-user@ip-10-0-0-135 Retail_Project]$
```

Step 3. Setup spark-streaming.py file in edit mode.

- a) Initialize environment variable for java home, spark home and pylib. Also import spark session and required pyspark sql libraries.
- b) Create Utility functions for calculating total value of single order which will be used in UDF. Here if order_type is "Order" then it should be 1 otherwise -1 for return order type.
- c) Then calculating unit price of each product with unit quantity. At the end returning the total order value from cost and type_multiplier.
- d) Again, in next utility functions for calculating total number of items present in single order which will be used in UDF. Here we are calculating total_item_count with summing up quantity across all the product and returning total_count.
- e) Again, create utility function for Is_Order and Is_Return attributes. In **Is_Order**, if order type is "Order" then return 1 otherwise 0. And for **Is_Return**, if order_type is "return" then return 1 otherwise return 0.
- f) Defining main function which start execution of UDF.
- g) Initialize spark session, setting appName as retailOrderAnalyzer and log_level as ERROR - we can set Warning also but this is good practice to set ERROR for production environment.
- h) Next, we need to read Input from Kafka so we have to provide bootstrap_server name with port, offsets and subscribed topic_name.
- i) Define Schema of product Order for invoice_no, timestamp, type, country, items, SKU, title, unit_price and quantity columns for reading data from Kafka.
- j) Define the 4 UDFs with the utility functions as these are required for preprocessing the order of 4 extra attributes total_cost, total_items, is_order and is_return.
- k) Next to write summarized input values in console. Writing in append mode on the console. Truncate mode is false as we don't want it to truncated. Returning data with processingTime = 1 minute interval and then start the job.
- l) Most important part is to Calculating KPIs for time and time-country based on defined schema. In both of the KPIs, for Tumbling window, keeping time as 1 minute, delay threshold keeping the same 1 minute for watermark. Running aggregation sum of total_cost, average of total_cost which is helping in calculating average_transaction_size. Count on Invoice_no which will give Order_Per_Minute count. Average of Is_return which will give rate_of_return.

- m) Calculate time and country based KPIs which is same like above but here we are getting time as well as country based KPIs data.
- n) Then, setting up in json format for time based KPIs in append mode, no need to truncate and store these json files in folder - **/tmp/op1** and checkpoint location is **/tmp/cp1**. And setting up in json format for time-and-country based KPIs in append mode, no need to truncate and store these files in folder - **/tmp/op2** and checkpoint location is **/tmp/cp2**.
- o) At last using awaitTermination method just wait for the termination of signal from User. And Save the file using command - **!wq**.

Step 4. Execute export command - **export SPARK_KAFKA_VERSION=0.10**

Step 5. Then execute spark2-submit command to execute the spark-streaming.py file. Also add host and port details along with Console_output file. Check below screenshot for reference.

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

```
ec2-user@ip-10-0-0-135:~/Retail Project
[ec2-user@ip-10-0-0-135 ~]$ export SPARK_KAFKA_VERSION=0.10
[ec2-user@ip-10-0-0-135 ~]$ 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
21/02/06 13:52:12 INFO spark.SparkContext: Running Spark version 2.3.0.cloudera2
21/02/06 13:52:12 INFO spark.SparkContext: Submitted application: retailOrderAnalyzer
21/02/06 13:52:13 INFO spark.SecurityManager: Changing view acls to: ec2-user
21/02/06 13:52:13 INFO spark.SecurityManager: Changing modify acls to: ec2-user
21/02/06 13:52:13 INFO spark.SecurityManager: Changing view acls groups to:
21/02/06 13:52:13 INFO spark.SecurityManager: Changing modify acls groups to:
21/02/06 13:52:13 INFO spark.SecurityManager: SecurityManager: authentication disabled; ui acls disabled; users with view permissions: Set(ec2-user); groups with view permissions: Set()
21/02/06 13:52:13 INFO util.Utils: Successfully started service 'sparkDriver' on port 34823.
21/02/06 13:52:13 INFO spark.SparkEnv: Registering MapOutputTracker
21/02/06 13:52:13 INFO spark.SparkEnv: Registering BlockManagerMaster
21/02/06 13:52:13 INFO storage.BlockManagerMasterEndpoint: Using org.apache.spark.storage.DefaultTopologyMapper for getting topology information
21/02/06 13:52:13 INFO storage.BlockManagerMasterEndpoint: BlockManagerMasterEndpoint up
21/02/06 13:52:13 INFO storage.DiskBlockManager: Created local directory at /tmp/blockmgr-11f87ddc-4dab-4c68-b597-0c2dcbcle7a2
21/02/06 13:52:13 INFO memory.MemoryStore: MemoryStore started with capacity 366.3 MB
21/02/06 13:52:13 INFO spark.SparkEnv: Registering OutputCommitCoordinator
21/02/06 13:52:13 INFO util.log: Logging initialized @4548ms
21/02/06 13:52:13 INFO server.Server: jetty-9.3.z-SNAPSHOT
21/02/06 13:52:13 INFO server.Server: Started @4674ms
21/02/06 13:52:13 INFO server.AbstractConnector: Started ServerConnector@3fe91be7[HTTP/1.1,(http://1.1.1.1:0.0.0.0:4040)]
21/02/06 13:52:13 INFO util.Utils: Successfully started service 'SparkUI' on port 4040.
21/02/06 13:52:13 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@4717fcb6(/jobs,null,AVAILABLE,@Spark)
21/02/06 13:52:13 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@6bad9718(/jobs/json,null,AVAILABLE,@Spark)
21/02/06 13:52:13 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@f69745e(/jobs/job,null,AVAILABLE,@Spark)
21/02/06 13:52:13 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@6b8f24d5(/jobs/job/json,null,AVAILABLE,@Spark)
21/02/06 13:52:13 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@5825fbb1(/stages,null,AVAILABLE,@Spark)
21/02/06 13:52:13 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@2873d628(/stages/json,null,AVAILABLE,@Spark)
21/02/06 13:52:13 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@77474905(/stages/stage,null,AVAILABLE,@Spark)
21/02/06 13:52:13 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@12dc71ce(/stages/stage/json,null,AVAILABLE,@Spark)
21/02/06 13:52:13 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@6134d8d(/stages/pool,null,AVAILABLE,@Spark)
21/02/06 13:52:13 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@5823eb35(/stages/pool/json,null,AVAILABLE,@Spark)
21/02/06 13:52:14 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@37d21bdc(/storage,null,AVAILABLE,@Spark)
21/02/06 13:52:14 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@5c2da405(/storage/json,null,AVAILABLE,@Spark)
21/02/06 13:52:14 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@5ab60292(/storage/rdd,null,AVAILABLE,@Spark)
21/02/06 13:52:14 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@74f5e6b3(/storage/rdd/json,null,AVAILABLE,@Spark)
21/02/06 13:52:14 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@6c24324(/environment,null,AVAILABLE,@Spark)
21/02/06 13:52:14 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@59c23e4a(/environment/json,null,AVAILABLE,@Spark)
21/02/06 13:52:14 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@62e332e5(/executors,null,AVAILABLE,@Spark)
21/02/06 13:52:14 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@640ca34(/executors/json,null,AVAILABLE,@Spark)
21/02/06 13:52:14 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@52e7b7fb(/executors/threadDump,null,AVAILABLE,@Spark)
21/02/06 13:52:14 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@58488e42(/executors/threadDump/json,null,AVAILABLE,@Spark)
21/02/06 13:52:14 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@58992907(/static,null,AVAILABLE,@Spark)
21/02/06 13:52:14 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@47616233(/,null,AVAILABLE,@Spark)
21/02/06 13:52:14 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@3bd266ed(/api,null,AVAILABLE,@Spark)
21/02/06 13:52:14 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@11c93cc(/jobs/job/kill,null,AVAILABLE,@Spark)
21/02/06 13:52:14 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@56b54ef5(/stages/stage/kill,null,AVAILABLE,@Spark)
21/02/06 13:52:14 INFO ui.SparkUI: Bound SparkUI to 0.0.0.0, and started at http://ip-10-0-0-135.ec2.internal:4040
21/02/06 13:52:14 INFO util.Utils: Using initial executors = 0, max of spark.dynamicAllocation.initialExecutors, spark.dynamicAllocation.minExecutors and spark.executor.instances
```

Here I have attached screenshot first 20 batches which are generating output. Here uses tail command to check batches of output file in another session.

tail -100f console_output

```
ec2-user@ip-10-0-0-135:~/Retail_Project
[ec2-user@ip-10-0-0-135 ~]$ cd --
[ec2-user@ip-10-0-0-135 ~]$ pwd
/home/ec2-user
[ec2-user@ip-10-0-0-135 ~]$ ls
airflow                               MapReduce_assignment                 Retail_Project
airflow-codes                         movie_ratings                       spark-sql-kafka-0-10_2.11-2.3.0.jar
assignment                           python3-3.6.8-13.el7.x86_64.rpm      spark-sql-kafka-0-10_2.11-2.3.0.jar.1
coding-labs                          python3-devel-3.6.8-13.el7.x86_64.rpm tmp
epel-release-latest-7.noarch.rpm      RetailDataAnalysis.py               u.user
final.py                             retail_data_project.py
install-airflow.sh                   retailDataProject.py
[ec2-user@ip-10-0-0-135 ~]$ cd Retail_Project/
[ec2-user@ip-10-0-0-135 Retail_Project]$ ls
Console-output  spark-sql-kafka-0-10_2.11-2.3.0.jar  spark-streaming.py  tmp
[ec2-user@ip-10-0-0-135 Retail_Project]$ tail -100f console_output
tail: cannot open 'console_output' for reading: No such file or directory
tail: no files remaining
[ec2-user@ip-10-0-0-135 Retail_Project]$ tail -100f Console-output
-----
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_order|is_return|
+-----+-----+-----+-----+-----+-----+
|154132542762167|United Kingdom|2021-02-06 13:52:37|73.6|152|1|0|
|154132542762168|United Kingdom|2021-02-06 13:52:40|45.449999999999996|25|1|0|
|154132542762169|United Kingdom|2021-02-06 13:52:41|46.9|17|1|0|
|154132542762170|United Kingdom|2021-02-06 13:52:41|65.320000000000001|8|1|0|
|154132542762171|United Kingdom|2021-02-06 13:52:44|100.30999999999999|13|1|0|
|154132542762172|United Kingdom|2021-02-06 13:52:48|9.56|5|1|0|
|154132542762173|United Kingdom|2021-02-06 13:52:48|66.18|34|1|0|
|154132542762174|United Kingdom|2021-02-06 13:52:54|26.419999999999998|35|1|0|
|154132542762175|United Kingdom|2021-02-06 13:53:01|85.35|17|1|0|
|154132542762176|United Kingdom|2021-02-06 13:53:04|9.499999999999998|13|1|0|
|154132542762177|Germany|2021-02-06 13:53:14|17.82|30|1|0|
|154132542762178|United Kingdom|2021-02-06 13:53:17|50.870000000000005|29|1|0|
|154132542762179|United Kingdom|2021-02-06 13:53:18|7.15|3|1|0|
|154132542762180|United Kingdom|2021-02-06 13:53:21|93.64999999999999|15|1|0|
|154132542762181|United Kingdom|2021-02-06 13:53:25|20.18|26|1|0|
+-----+-----+-----+-----+-----+-----+
```

ls -lC -user@ip-10-0-0-133 ~/netdata/PROJECT

```
|154132542762171|United Kingdom|2021-02-06 13:52:44|100.30999999999999|13|1|1|0|
|154132542762172|United Kingdom|2021-02-06 13:52:48|9.56|5|1|1|0|
|154132542762173|United Kingdom|2021-02-06 13:52:48|66.18|34|1|1|0|
|154132542762174|United Kingdom|2021-02-06 13:52:54|26.419999999999998|35|1|1|0|
|154132542762175|United Kingdom|2021-02-06 13:53:01|85.35|17|1|1|0|
|154132542762176|United Kingdom|2021-02-06 13:53:04|9.499999999999998|13|1|1|0|
|154132542762177|Germany|2021-02-06 13:53:14|17.82|30|1|1|0|
|154132542762178|United Kingdom|2021-02-06 13:53:17|50.870000000000005|29|1|1|0|
|154132542762179|United Kingdom|2021-02-06 13:53:18|7.15|3|1|1|0|
|154132542762180|United Kingdom|2021-02-06 13:53:21|93.64999999999999|15|1|1|0|
|154132542762181|United Kingdom|2021-02-06 13:53:25|20.18|26|1|1|0|
+-----+-----+-----+-----+-----+-----+-----+-----+
-----
```

Batch: 2

```
+-----+-----+-----+-----+-----+-----+-----+-----+
|invoice_no|country|timestamp|total_cost|total_items|is_order|is_return|
+-----+-----+-----+-----+-----+-----+-----+-----+
|154132542762182|United Kingdom|2021-02-06 13:53:28|14.76|6|1|1|0|
|154132542762183|United Kingdom|2021-02-06 13:53:30|18.380000000000003|6|1|1|0|
|154132542762184|United Kingdom|2021-02-06 13:53:32|39.45|9|1|1|0|
|154132542762185|United Kingdom|2021-02-06 13:53:32|75.81|31|1|1|0|
|154132542762186|United Kingdom|2021-02-06 13:53:34|17.1|5|1|1|0|
|154132542762187|United Kingdom|2021-02-06 13:53:36|16.5|2|1|1|0|
|154132542762188|United Kingdom|2021-02-06 13:53:38|4.2|2|1|1|0|
|154132542762189|EIRE|2021-02-06 13:53:42|6.35|3|1|1|0|
|154132542762190|United Kingdom|2021-02-06 13:53:45|21.849999999999998|32|1|1|0|
|154132542762191|Switzerland|2021-02-06 13:53:52|15.66|15|1|1|0|
|154132542762192|United Kingdom|2021-02-06 13:53:52|-333.55|141|0|1|1|
|154132542762193|United Kingdom|2021-02-06 13:54:00|1.7|2|1|1|0|
|154132542762194|United Kingdom|2021-02-06 13:54:02|33.16|4|1|1|0|
|154132542762195|United Kingdom|2021-02-06 13:54:10|11.03|13|1|1|0|
|154132542762196|France|2021-02-06 13:54:17|42.11|20|1|1|0|
+-----+-----+-----+-----+-----+-----+-----+-----+
-----
```

Batch: 3

```
+-----+-----+-----+-----+-----+-----+-----+-----+
|invoice_no|country|timestamp|total_cost|total_items|is_order|is_return|
+-----+-----+-----+-----+-----+-----+-----+-----+
|154132542762197|United Kingdom|2021-02-06 13:54:28|2.55|1|1|1|0|
|154132542762198|United Kingdom|2021-02-06 13:54:30|0.75|1|1|1|0|
|154132542762199|United Kingdom|2021-02-06 13:55:04|77.54|16|1|1|0|
|154132542762200|United Kingdom|2021-02-06 13:55:13|85.13|70|1|1|0|
|154132542762201|United Kingdom|2021-02-06 13:55:14|46.18|26|1|1|0|
+-----+-----+-----+-----+-----+-----+-----+-----+
-----
```

ec2-user@ip-10-0-0-135:~/Retail_Project

Batch: 4

invoice_no	country	timestamp	total_cost	total_items	is_order	is_return
154132542762202	United Kingdom	2021-02-06 13:55:31	34.510000000000005	17	1	0
154132542762203	EIRE	2021-02-06 13:55:32	19.799999999999997	12	1	0
154132542762204	France	2021-02-06 13:55:33	-53.95	17	0	1
154132542762205	United Kingdom	2021-02-06 13:55:49	22.729999999999997	21	1	0
154132542762206	United Kingdom	2021-02-06 13:55:54	17.0	2	1	0
154132542762207	United Kingdom	2021-02-06 13:56:24	-82.44	16	0	1

Batch: 5

invoice_no	country	timestamp	total_cost	total_items	is_order	is_return
154132542762208	Spain	2021-02-06 13:56:28	1.63	1	1	0
154132542762209	United Kingdom	2021-02-06 13:56:44	5.04	24	1	0
154132542762210	United Kingdom	2021-02-06 13:56:52	118.77	29	1	0
154132542762211	United Kingdom	2021-02-06 13:56:58	44.72	26	1	0
154132542762212	United Kingdom	2021-02-06 13:57:06	252.36	63	1	0
154132542762213	United Kingdom	2021-02-06 13:57:10	19.56	12	1	0
154132542762214	United Kingdom	2021-02-06 13:57:20	254.42	95	1	0

Batch: 6

invoice_no	country	timestamp	total_cost	total_items	is_order	is_return
154132542762215	United Kingdom	2021-02-06 13:57:23	96.47	69	1	0
154132542762216	United Kingdom	2021-02-06 13:57:23	63.81999999999999	24	1	0
154132542762217	United Kingdom	2021-02-06 13:57:23	5.2	2	1	0
154132542762218	United Kingdom	2021-02-06 13:57:25	203.25	43	1	0
154132542762219	United Kingdom	2021-02-06 13:57:26	29.25	7	1	0
154132542762220	United Kingdom	2021-02-06 13:57:26	23.549999999999997	15	1	0
154132542762221	France	2021-02-06 13:57:33	13.98	37	1	0
154132542762222	United Kingdom	2021-02-06 13:57:45	94.35	13	1	0
154132542762223	United Kingdom	2021-02-06 13:57:49	23.6	8	1	0
154132542762224	United Kingdom	2021-02-06 13:57:55	101.03999999999999	52	1	0
154132542762225	United Kingdom	2021-02-06 13:57:56	2.95	1	1	0
154132542762226	United Kingdom	2021-02-06 13:58:07	8.7	4	1	0
154132542762227	United Kingdom	2021-02-06 13:58:18	3.3	2	1	0
154132542762228	United Kingdom	2021-02-06 13:58:21	118.76	26	1	0

ec2-user@ip-10-0-0-135:~/Retail_Project

```
|154132542762227|United Kingdom|2021-02-06 13:58:18|3.3|2|1|0|
|154132542762228|United Kingdom|2021-02-06 13:58:21|118.76|26|1|0|
+-----+-----+-----+-----+-----+-----+-----+
```

Batch: 7

```
+-----+-----+-----+-----+-----+-----+-----+
|invoice_no|country|timestamp|total_cost|total_items|is_order|is_return|
+-----+-----+-----+-----+-----+-----+-----+
|154132542762229|United Kingdom|2021-02-06 13:58:26|4.62|6|1|0|
|154132542762230|United Kingdom|2021-02-06 13:58:48|1.25|1|1|0|
|154132542762231|United Kingdom|2021-02-06 13:58:52|1.66|3|1|0|
|154132542762232|United Kingdom|2021-02-06 13:59:10|90.47|37|1|0|
+-----+-----+-----+-----+-----+-----+-----+
```

Batch: 8

```
+-----+-----+-----+-----+-----+-----+-----+
|invoice_no|country|timestamp|total_cost|total_items|is_order|is_return|
+-----+-----+-----+-----+-----+-----+-----+
|154132542762233|United Kingdom|2021-02-06 13:59:25|70.14000000000001|18|1|0|
|154132542762234|United Kingdom|2021-02-06 13:59:30|51.269999999999996|9|1|0|
|154132542762235|United Kingdom|2021-02-06 13:59:48|45.82|14|1|0|
|154132542762236|United Kingdom|2021-02-06 13:59:51|34.8|8|1|0|
|154132542762237|United Kingdom|2021-02-06 14:00:02|3.75|1|1|0|
|154132542762238|Australia|2021-02-06 14:00:05|92.37|43|1|0|
|154132542762239|United Kingdom|2021-02-06 14:00:12|51.809999999999995|21|1|0|
|154132542762240|United Kingdom|2021-02-06 14:00:17|35.4|36|1|0|
|154132542762241|United Kingdom|2021-02-06 14:00:17|64.88|25|1|0|
|154132542762242|United Kingdom|2021-02-06 14:00:20|16.5|10|1|0|
+-----+-----+-----+-----+-----+-----+-----+
```

Batch: 9

```
+-----+-----+-----+-----+-----+-----+-----+
|invoice_no|country|timestamp|total_cost|total_items|is_order|is_return|
+-----+-----+-----+-----+-----+-----+-----+
|154132542762243|United Kingdom|2021-02-06 14:00:30|107.17000000000002|63|1|0|
|154132542762244|United Kingdom|2021-02-06 14:00:31|7.5|6|1|0|
|154132542762245|United Kingdom|2021-02-06 14:00:48|30.42|25|1|0|
|154132542762246|United Kingdom|2021-02-06 14:00:54|59.519999999999996|12|1|0|
|154132542762247|United Kingdom|2021-02-06 14:01:06|76.0|16|1|0|
|154132542762248|United Kingdom|2021-02-06 14:01:07|273.89|131|1|0|
|154132542762249|United Kingdom|2021-02-06 14:01:11|107.63999999999999|23|1|0|
+-----+-----+-----+-----+-----+-----+-----+
```


ec2-user@ip-10-0-0-135:~/Retail_Project

```
|154132542762244|United Kingdom|2021-02-06 14:00:31|7.5|6|1|0|
|154132542762245|United Kingdom|2021-02-06 14:00:48|30.42|25|1|0|
|154132542762246|United Kingdom|2021-02-06 14:00:54|59.519999999999996|12|1|0|
|154132542762247|United Kingdom|2021-02-06 14:01:06|76.0|16|1|0|
|154132542762248|United Kingdom|2021-02-06 14:01:07|273.89|131|1|0|
|154132542762249|United Kingdom|2021-02-06 14:01:11|107.63999999999999|23|1|0|
+-----+-----+-----+-----+-----+-----+-----+-----+
|
```

Batch: 10

```
+-----+-----+-----+-----+-----+-----+-----+-----+
|invoice_no|country|timestamp|total_cost|total_items|is_order|is_return|
+-----+-----+-----+-----+-----+-----+-----+-----+
|154132542762250|United Kingdom|2021-02-06 14:01:23|48.0|14|1|0|
|154132542762251|Germany|2021-02-06 14:01:30|91.66|20|1|0|
|154132542762252|United Kingdom|2021-02-06 14:01:40|6.24|3|1|0|
|154132542762253|United Kingdom|2021-02-06 14:01:42|143.100000000000002|86|1|0|
|154132542762254|United Kingdom|2021-02-06 14:01:43|71.929999999999999|31|1|0|
|154132542762255|United Kingdom|2021-02-06 14:01:55|74.0|32|1|0|
|154132542762256|United Kingdom|2021-02-06 14:02:00|31.4900000000000002|9|1|0|
|154132542762257|United Kingdom|2021-02-06 14:02:04|10.5|25|1|0|
|154132542762258|United Kingdom|2021-02-06 14:02:07|23.93|15|1|0|
|154132542762259|United Kingdom|2021-02-06 14:02:09|25.85|11|1|0|
|154132542762260|United Kingdom|2021-02-06 14:02:13|299.3|104|1|0|
|154132542762261|United Kingdom|2021-02-06 14:02:16|183.89999999999998|208|1|0|
|154132542762262|United Kingdom|2021-02-06 14:02:18|30.869999999999997|20|1|0|
+-----+-----+-----+-----+-----+-----+-----+-----+
|
```

Batch: 11

```
+-----+-----+-----+-----+-----+-----+-----+-----+
|invoice_no|country|timestamp|total_cost|total_items|is_order|is_return|
+-----+-----+-----+-----+-----+-----+-----+-----+
|154132542762263|United Kingdom|2021-02-06 14:02:24|30.0|24|1|0|
|154132542762264|United Kingdom|2021-02-06 14:02:28|39.349999999999994|13|1|0|
|154132542762265|United Kingdom|2021-02-06 14:02:28|19.56|12|1|0|
|154132542762266|United Kingdom|2021-02-06 14:02:41|479.42000000000001|58|1|0|
|154132542762267|United Kingdom|2021-02-06 14:02:44|38.9|6|1|0|
|154132542762268|United Kingdom|2021-02-06 14:02:52|15.0|4|1|0|
|154132542762269|United Kingdom|2021-02-06 14:02:58|7.5|1|1|0|
|154132542762270|Germany|2021-02-06 14:02:59|80.69999999999999|26|1|0|
|154132542762271|United Kingdom|2021-02-06 14:03:00|118.320000000000001|45|1|0|
|154132542762272|United Kingdom|2021-02-06 14:03:00|269.26|132|1|0|
|154132542762273|United Kingdom|2021-02-06 14:03:07|44.94|22|1|0|
|154132542762274|United Kingdom|2021-02-06 14:03:08|32.37|17|1|0|
+-----+-----+-----+-----+-----+-----+-----+-----+
|
```

ec2-user@ip-10-0-0-135:~/Retail_Project

Batch: 12

invoice_no	country	timestamp	total_cost	total_items	is_order	is_return
154132542762275	United Kingdom	2021-02-06 14:03:18	16.29	23	1	0
154132542762276	United Kingdom	2021-02-06 14:03:23	-46.35	19	0	1
154132542762277	United Kingdom	2021-02-06 14:03:27	1.52	8	1	0
154132542762278	United Kingdom	2021-02-06 14:03:27	41.34	5	1	0
154132542762279	EIRE	2021-02-06 14:03:30	45.55	11	1	0
154132542762280	United Kingdom	2021-02-06 14:03:35	19.8	4	1	0
154132542762281	United Kingdom	2021-02-06 14:03:35	29.700000000000003	6	1	0
154132542762282	United Kingdom	2021-02-06 14:03:40	19.01	5	1	0
154132542762283	United Kingdom	2021-02-06 14:03:49	39.609999999999999	17	1	0
154132542762284	United Kingdom	2021-02-06 14:03:51	49.93	10	1	0
154132542762285	United Kingdom	2021-02-06 14:03:54	19.049999999999997	9	1	0
154132542762286	United Kingdom	2021-02-06 14:03:59	16.64	8	1	0
154132542762287	United Kingdom	2021-02-06 14:04:04	5.45	3	1	0
154132542762288	United Kingdom	2021-02-06 14:04:12	26.49	15	1	0

Batch: 13

invoice_no	country	timestamp	total_cost	total_items	is_order	is_return
154132542762289	United Kingdom	2021-02-06 14:04:18	25.5	3	1	0
154132542762290	United Kingdom	2021-02-06 14:04:27	16.65	7	1	0
154132542762291	United Kingdom	2021-02-06 14:04:30	53.370000000000005	31	1	0
154132542762292	United Kingdom	2021-02-06 14:04:30	9.96	4	1	0
154132542762293	United Kingdom	2021-02-06 14:04:51	40.5	49	1	0
154132542762294	United Kingdom	2021-02-06 14:04:56	32.010000000000005	21	1	0

Batch: 14

invoice_no	country	timestamp	total_cost	total_items	is_order	is_return
154132542762295	United Kingdom	2021-02-06 14:05:19	311.09999999999997	146	1	0
154132542762296	United Kingdom	2021-02-06 14:05:34	29.759999999999998	6	1	0
154132542762297	United Kingdom	2021-02-06 14:05:36	98.4	40	1	0
154132542762298	United Kingdom	2021-02-06 14:05:46	8.5	2	1	0
154132542762299	United Kingdom	2021-02-06 14:06:03	40.269999999999996	19	1	0
154132542762300	United Kingdom	2021-02-06 14:06:06	117.63000000000001	19	1	0
154132542762301	United Kingdom	2021-02-06 14:06:12	50.58	68	1	0

ec2-user@ip-10-0-0-135:~/Retail_Project

Batch: 15

invoice_no	country	timestamp	total_cost	total_items	is_order	is_return
154132542762303	United Kingdom	2021-02-06 14:06:15	51.43	13	1	0
154132542762304	United Kingdom	2021-02-06 14:06:19	96.84	34	1	0
154132542762305	United Kingdom	2021-02-06 14:06:35	271.6	130	1	0
154132542762306	United Kingdom	2021-02-06 14:06:37	11.59	3	1	0
154132542762307	United Kingdom	2021-02-06 14:06:39	574.23	160	1	0
154132542762308	United Kingdom	2021-02-06 14:06:40	97.98	47	1	0
154132542762309	United Kingdom	2021-02-06 14:06:46	150.63	56	1	0
154132542762310	United Kingdom	2021-02-06 14:06:46	74.81	47	1	0
154132542762311	United Kingdom	2021-02-06 14:07:03	45.83	29	1	0
154132542762312	United Kingdom	2021-02-06 14:07:04	74.85999999999999	23	1	0
154132542762313	United Kingdom	2021-02-06 14:07:09	6.64	8	1	0

Batch: 16

invoice_no	country	timestamp	total_cost	total_items	is_order	is_return
154132542762314	United Kingdom	2021-02-06 14:07:17	55.32	16	1	0
154132542762315	United Kingdom	2021-02-06 14:07:28	351.05	106	1	0
154132542762316	United Kingdom	2021-02-06 14:07:28	59.35	13	1	0
154132542762317	United Kingdom	2021-02-06 14:07:40	62.24	26	1	0
154132542762318	United Kingdom	2021-02-06 14:07:41	33.9	2	1	0
154132542762319	United Kingdom	2021-02-06 14:08:04	16.45	13	1	0
154132542762320	United Kingdom	2021-02-06 14:08:09	154.11999999999998	68	1	0
154132542762321	United Kingdom	2021-02-06 14:08:13	21.6	12	1	0

Batch: 17

invoice_no	country	timestamp	total_cost	total_items	is_order	is_return
154132542762322	United Kingdom	2021-02-06 14:08:13	1.7	2	1	0
154132542762323	United Kingdom	2021-02-06 14:08:16	189.51	51	1	0
154132542762324	United Kingdom	2021-02-06 14:08:20	45.239999999999995	84	1	0
154132542762325	United Kingdom	2021-02-06 14:08:24	7.75	3	1	0
154132542762326	United Kingdom	2021-02-06 14:08:29	189.24	190	1	0
154132542762327	United Kingdom	2021-02-06 14:08:30	9.56	2	1	0
154132542762328	United Kingdom	2021-02-06 14:08:31	72.09	175	1	0
154132542762329	United Kingdom	2021-02-06 14:08:36	27.25	15	1	0

ec2-user@ip-10-0-0-135:~/Retail_Project

Batch: 18

invoice_no	country	timestamp	total_cost	total_items	is_order	is_return
154132542762336	United Kingdom	2021-02-06 14:09:12	29.86	14	1	0
154132542762337	EIRE	2021-02-06 14:09:16	159.99	95	1	0
154132542762338	United Kingdom	2021-02-06 14:09:22	-45.45	61	0	1
154132542762339	United Kingdom	2021-02-06 14:09:23	87.84	32	1	0
154132542762340	United Kingdom	2021-02-06 14:09:26	154.07999999999998	62	1	0
154132542762341	Portugal	2021-02-06 14:09:30	19.230000000000004	16	1	0
154132542762342	Australia	2021-02-06 14:09:32	-355.88	156	0	1
154132542762343	United Kingdom	2021-02-06 14:09:43	108.22	26	1	0
154132542762344	United Kingdom	2021-02-06 14:09:48	180.27	88	1	0
154132542762345	United Kingdom	2021-02-06 14:09:50	49.18	34	1	0
154132542762346	United Kingdom	2021-02-06 14:09:55	32.760000000000005	18	1	0
154132542762347	United Kingdom	2021-02-06 14:09:57	44.54	16	1	0
154132542762348	United Kingdom	2021-02-06 14:09:58	11.7	6	1	0
154132542762349	United Kingdom	2021-02-06 14:10:00	304.92	220	1	0

Batch: 19

invoice_no	country	timestamp	total_cost	total_items	is_order	is_return
154132542762350	United Kingdom	2021-02-06 14:10:17	7.95	1	1	0
154132542762351	United Kingdom	2021-02-06 14:10:25	30.1	33	1	0
154132542762352	United Kingdom	2021-02-06 14:10:34	8.66	3	1	0
154132542762353	United Kingdom	2021-02-06 14:10:36	16.25	11	1	0
154132542762354	United Kingdom	2021-02-06 14:10:38	25.200000000000003	12	1	0
154132542762355	EIRE	2021-02-06 14:10:42	42.19	18	1	0
154132542762356	United Kingdom	2021-02-06 14:10:42	204.0	24	1	0
154132542762357	United Kingdom	2021-02-06 14:10:44	178.93	58	1	0
154132542762358	United Kingdom	2021-02-06 14:10:44	189.12	144	1	0
154132542762359	United Kingdom	2021-02-06 14:10:48	36.3	34	1	0
154132542762360	United Kingdom	2021-02-06 14:10:52	96.24999999999999	47	1	0
154132542762361	United Kingdom	2021-02-06 14:10:59	-30.48	12	0	1
154132542762362	United Kingdom	2021-02-06 14:11:03	3.36	8	1	0
154132542762363	United Kingdom	2021-02-06 14:11:09	-36.49	11	0	1

Batch: 20

Files in folder :

1) /tmp/op1/ Also read the record from json file at the end.

Hadoop fs -cat /tmp/op1/<json-file-name>

```
[ec2-user@ip-10-0-0-135 ~]$ cd Retail_Project/
[ec2-user@ip-10-0-0-135 Retail_Project]$ ls
Console-output  spark-sql-kafka-0-10  2.11-2.3.0.jar  spark-streaming.py  tmp
[ec2-user@ip-10-0-0-135 Retail_Project]$ hadoop fs -ls /tmp/op1
Found 29 items
drwxr-xr-x - ec2-user supergroup 0 2021-02-06 14:02 /tmp/op1/ spark metadata
-rw-r--r-- 3 ec2-user supergroup 0 2021-02-06 13:54 /tmp/op1/part-00000-0a83e125-a84a-43ed-bed0-8118977a3d5a-c000.json
-rw-r--r-- 3 ec2-user supergroup 0 2021-02-06 13:55 /tmp/op1/part-00000-13680399-5424-4474-b8aa-8e64c5cd9f3c-c000.json
-rw-r--r-- 3 ec2-user supergroup 0 2021-02-06 13:53 /tmp/op1/part-00000-1ae7b9c2-cf7f-42d9-b0b5-68de1a3d70b3-c000.json
-rw-r--r-- 3 ec2-user supergroup 0 2021-02-06 13:54 /tmp/op1/part-00000-24e6ef12-fd9c-4896-bb2b-2a51f7f13eb3-c000.json
-rw-r--r-- 3 ec2-user supergroup 0 2021-02-06 14:02 /tmp/op1/part-00000-387ebb45-ab31-481f-a26e-89573f88acb7-c000.json
-rw-r--r-- 3 ec2-user supergroup 0 2021-02-06 13:59 /tmp/op1/part-00000-64e54fe2-fb74-41c4-a082-d3397e07d3f9-c000.json
-rw-r--r-- 3 ec2-user supergroup 0 2021-02-06 13:47 /tmp/op1/part-00000-75f4e8f6-0255-4c6b-912a-7d572a07eac8-c000.json
-rw-r--r-- 3 ec2-user supergroup 0 2021-02-06 13:57 /tmp/op1/part-00000-908b9a50-2a99-4cba-8aec-20c935c23e09-c000.json
-rw-r--r-- 3 ec2-user supergroup 0 2021-02-06 14:01 /tmp/op1/part-00000-912bebe0-d192-46f7-a233-0c188ead01c-c000.json
-rw-r--r-- 3 ec2-user supergroup 0 2021-02-06 14:03 /tmp/op1/part-00000-b414ebab-6da2-4a86-8a9a-c0125bc1879f-c000.json
-rw-r--r-- 3 ec2-user supergroup 0 2021-02-06 13:50 /tmp/op1/part-00000-e68c631e-92c6-46c8-a14e-d2097d0b3f40-c000.json
-rw-r--r-- 3 ec2-user supergroup 0 2021-02-06 13:58 /tmp/op1/part-00000-e80f4373-ad3d-4c47-8eaf-1f2b43731110-c000.json
-rw-r--r-- 3 ec2-user supergroup 0 2021-02-06 13:49 /tmp/op1/part-00000-eba25967-bcd5-4225-bb6f-b17280446fef-c000.json
-rw-r--r-- 3 ec2-user supergroup 0 2021-02-06 13:48 /tmp/op1/part-00000-ef3199db-ec6e-4765-8d5a-eacc0a1fb584-c000.json
-rw-r--r-- 3 ec2-user supergroup 0 2021-02-06 14:00 /tmp/op1/part-00000-f12ef73f-817b-485b-83da-232744522078-c000.json
-rw-r--r-- 3 ec2-user supergroup 177 2021-02-06 13:50 /tmp/op1/part-00008-edf54801-176e-42d1-85a4-38ebc8fad3c3-c000.json
-rw-r--r-- 3 ec2-user supergroup 178 2021-02-06 13:54 /tmp/op1/part-00031-ff184248-e0c7-4579-a1bf-4bfb9b8175af-c000.json
-rw-r--r-- 3 ec2-user supergroup 176 2021-02-06 13:58 /tmp/op1/part-00047-cb48cb5d-fe8b-413f-a1d4-ed9a1b212e6c-c000.json
-rw-r--r-- 3 ec2-user supergroup 178 2021-02-06 13:53 /tmp/op1/part-00055-3e8d050a-65a5-4bbc-a382-c0639bda5b2e-c000.json
-rw-r--r-- 3 ec2-user supergroup 178 2021-02-06 13:57 /tmp/op1/part-00068-1ff4c7ad-96df-4a24-ac8b-d285b11aae53-c000.json
-rw-r--r-- 3 ec2-user supergroup 180 2021-02-06 13:55 /tmp/op1/part-00083-2aae73a7-7254-486f-9e55-c4a5ff939920-c000.json
-rw-r--r-- 3 ec2-user supergroup 177 2021-02-06 13:55 /tmp/op1/part-00106-dad8bf0c-6252-4e28-88ee-defc1fa4824a-c000.json
-rw-r--r-- 3 ec2-user supergroup 181 2021-02-06 13:55 /tmp/op1/part-00124-03e4d992-00d8-4da4-9d18-05c19b5dc566-c000.json
-rw-r--r-- 3 ec2-user supergroup 176 2021-02-06 14:00 /tmp/op1/part-00124-19437c04-1c2d-447e-99b9-e58d8cd40cc1-c000.json
-rw-r--r-- 3 ec2-user supergroup 178 2021-02-06 13:55 /tmp/op1/part-00130-aa5ab4e3-5199-4540-a6de-d90dc4ce2f81-c000.json
-rw-r--r-- 3 ec2-user supergroup 180 2021-02-06 14:01 /tmp/op1/part-00146-54c394b5-488e-43c2-84ec-b36bf5041a1-c000.json
-rw-r--r-- 3 ec2-user supergroup 177 2021-02-06 13:59 /tmp/op1/part-00167-01334173-59cf-4db3-827a-988dd7aa03c-c000.json
-rw-r--r-- 3 ec2-user supergroup 177 2021-02-06 14:02 /tmp/op1/part-00196-f295dc9a-0b1e-4380-8874-abd54e501377-c000.json
[ec2-user@ip-10-0-0-135 Retail_Project]$ hadoop fs -cat /tmp/op1/part-00196-f295dc9a-0b1e-4380-8874-abd54e501377-c000.json
{"window": {"start": "2021-02-06T13:58:00.000Z", "end": "2021-02-06T13:59:00.000Z"}, "OPM": 6, "total_sale_volume": "138.29", "average_transaction_size": "23.05", "rate_of_return": "0.00"}
```

To check list of files in hadoop → **hadoop fs -ls /tmp/op1**

```
ec2-user@ip-10-0-0-135:~/Retail_Project
[ec2-user@ip-10-0-0-135 Retail_Project]$ hadoop fs -ls /tmp/op1
Found 42 items
drwxr-xr-x - ec2-user supergroup 0 2021-02-06 14:09 /tmp/op1/ spark metadata
-rw-r--r-- 3 ec2-user supergroup 0 2021-02-06 13:54 /tmp/op1/part-00000-0a83e125-a84a-43ed-bed0-8118977a3d5a-c000.json
-rw-r--r-- 3 ec2-user supergroup 0 2021-02-06 13:55 /tmp/op1/part-00000-13680399-5424-4474-b8aa-8e64c5cd9f3c-c000.json
-rw-r--r-- 3 ec2-user supergroup 0 2021-02-06 13:53 /tmp/op1/part-00000-1ae7b9c2-cf7f-42d9-b0b5-68de1a3d70b3-c000.json
-rw-r--r-- 3 ec2-user supergroup 0 2021-02-06 13:54 /tmp/op1/part-00000-24e6ef12-fd9c-4896-bb2b-2a51f7f13eb3-c000.json
-rw-r--r-- 3 ec2-user supergroup 0 2021-02-06 14:08 /tmp/op1/part-00000-2ec158bc-8cbf-4fca-8545-98317278b10b-c000.json
-rw-r--r-- 3 ec2-user supergroup 0 2021-02-06 14:02 /tmp/op1/part-00000-387ebb45-ab31-481f-a26e-89573f88acb7-c000.json
-rw-r--r-- 3 ec2-user supergroup 0 2021-02-06 14:05 /tmp/op1/part-00000-3eaf89e4-4d14-40b9-a6cc-9ec00bd28ac4-c000.json
-rw-r--r-- 3 ec2-user supergroup 0 2021-02-06 14:04 /tmp/op1/part-00000-4f42291f-8fd9-4727-9ea1-45d8bdccce4b-c000.json
-rw-r--r-- 3 ec2-user supergroup 0 2021-02-06 14:07 /tmp/op1/part-00000-5ece5b3c-0737-42db-a0e8-ff8ecf7b94f2-c000.json
-rw-r--r-- 3 ec2-user supergroup 0 2021-02-06 13:59 /tmp/op1/part-00000-64e54fe2-fb74-41c4-a082-d3397e07d3f9-c000.json
-rw-r--r-- 3 ec2-user supergroup 0 2021-02-06 13:47 /tmp/op1/part-00000-75f4e8f6-0255-4c6b-912a-7d572a07eac8-c000.json
-rw-r--r-- 3 ec2-user supergroup 0 2021-02-06 14:06 /tmp/op1/part-00000-7fde6a76-827b-423c-b8ce-116e8915cae3-c000.json
-rw-r--r-- 3 ec2-user supergroup 0 2021-02-06 13:57 /tmp/op1/part-00000-908b9a50-2a99-4cba-8aec-20c935c23e09-c000.json
-rw-r--r-- 3 ec2-user supergroup 0 2021-02-06 14:01 /tmp/op1/part-00000-912bebe0-d192-46f7-a233-0c188ead01c-c000.json
-rw-r--r-- 3 ec2-user supergroup 0 2021-02-06 14:09 /tmp/op1/part-00000-aaef1a52-5dde-4591-a926-4cc2cc99f179-c000.json
-rw-r--r-- 3 ec2-user supergroup 0 2021-02-06 14:03 /tmp/op1/part-00000-b414ebab-6da2-4a86-8a9a-c0125bc1879f-c000.json
-rw-r--r-- 3 ec2-user supergroup 0 2021-02-06 13:50 /tmp/op1/part-00000-e68c631e-92c6-46c8-a14e-d2097d0b3f40-c000.json
-rw-r--r-- 3 ec2-user supergroup 0 2021-02-06 13:58 /tmp/op1/part-00000-e80f4373-ad3d-4c47-8eaf-1f2b43731110-c000.json
-rw-r--r-- 3 ec2-user supergroup 0 2021-02-06 13:49 /tmp/op1/part-00000-eba25967-bcd5-4225-bb6f-b17280446fef-c000.json
-rw-r--r-- 3 ec2-user supergroup 0 2021-02-06 13:48 /tmp/op1/part-00000-ef3199db-ec6e-4765-8d5a-eacc0a1fb584-c000.json
-rw-r--r-- 3 ec2-user supergroup 0 2021-02-06 14:00 /tmp/op1/part-00000-f12ef73f-817b-485b-83da-232744522078-c000.json
-rw-r--r-- 3 ec2-user supergroup 180 2021-02-06 14:06 /tmp/op1/part-00006-6ae8b734-31d0-434d-a051-b7e1561720cf-c000.json
-rw-r--r-- 3 ec2-user supergroup 177 2021-02-06 13:50 /tmp/op1/part-00008-edf54801-176e-42d1-85a4-38ebc8fad3c3-c000.json
-rw-r--r-- 3 ec2-user supergroup 178 2021-02-06 14:04 /tmp/op1/part-00025-acaf532c-0605-42f1-a5ca-1e0168848404-c000.json
-rw-r--r-- 3 ec2-user supergroup 178 2021-02-06 13:54 /tmp/op1/part-00031-ff184248-e0c7-4579-a1bf-4bfb9b8175af-c000.json
-rw-r--r-- 3 ec2-user supergroup 176 2021-02-06 13:58 /tmp/op1/part-00047-cb48cb5d-fe8b-413f-a1d4-ed9a1b212e6c-c000.json
-rw-r--r-- 3 ec2-user supergroup 178 2021-02-06 13:53 /tmp/op1/part-00055-3e8d050a-65a5-4bbc-a382-c0639bda5b2e-c000.json
-rw-r--r-- 3 ec2-user supergroup 178 2021-02-06 13:57 /tmp/op1/part-00068-1ff4c7ad-96df-4a24-ac8b-d285b11aae53-c000.json
-rw-r--r-- 3 ec2-user supergroup 180 2021-02-06 13:55 /tmp/op1/part-00083-2aae73a7-7254-486f-9e55-c4a5ff939920-c000.json
-rw-r--r-- 3 ec2-user supergroup 177 2021-02-06 14:05 /tmp/op1/part-00090-2007fcd7-f80b-4005-84a1-51617ef10541-c000.json
-rw-r--r-- 3 ec2-user supergroup 177 2021-02-06 14:03 /tmp/op1/part-00097-4f1444d7-f69e-4014-96ef-b2b18e2ebcf8-c000.json
-rw-r--r-- 3 ec2-user supergroup 177 2021-02-06 13:55 /tmp/op1/part-00106-dad8bf0c-6252-4e28-88ee-defc1fa4824a-c000.json
-rw-r--r-- 3 ec2-user supergroup 181 2021-02-06 13:55 /tmp/op1/part-00124-03e4d992-00d8-4da4-9d18-05c19b5dc566-c000.json
-rw-r--r-- 3 ec2-user supergroup 176 2021-02-06 14:00 /tmp/op1/part-00124-19437c04-1c2d-447e-99b9-e58d8cd40cc1-c000.json
-rw-r--r-- 3 ec2-user supergroup 178 2021-02-06 13:55 /tmp/op1/part-00130-aa5ab4e3-5199-4540-a6de-d90dc4ce2f81-c000.json
-rw-r--r-- 3 ec2-user supergroup 178 2021-02-06 14:09 /tmp/op1/part-00132-a3b296a8-3497-4802-bf6c-5c95c30cd2dc-c000.json
-rw-r--r-- 3 ec2-user supergroup 180 2021-02-06 14:01 /tmp/op1/part-00146-54c394b5-488e-43c2-84ec-b36bf5041a1-c000.json
-rw-r--r-- 3 ec2-user supergroup 178 2021-02-06 14:08 /tmp/op1/part-00151-9dca264e-7773-4c08-9fef-99a876992667-c000.json
-rw-r--r-- 3 ec2-user supergroup 177 2021-02-06 14:08 /tmp/op1/part-00164-2765e20f-afc2-4495-95da-017cda3c134d-c000.json
-rw-r--r-- 3 ec2-user supergroup 177 2021-02-06 13:59 /tmp/op1/part-00167-01334173-59cf-4db3-827a-988dd7aa03c-c000.json
-rw-r--r-- 3 ec2-user supergroup 177 2021-02-06 14:02 /tmp/op1/part-00196-f295dc9a-0b1e-4380-8874-abd54e501377-c000.json
[ec2-user@ip-10-0-0-135 Retail_Project]$ hadoop fs -ls /tmp/op2
Found 60 items
```

2) /tmp/op2/

```
ec2-user@ip-10-0-0-135:~/Retail_Project
{"window":{"start":"2021-02-06T13:58:00.000Z","end":"2021-02-06T13:59:00.000Z"},"OPM":6,"total_sale_volume":"138.29","average_trans":1}
[ec2-user@ip-10-0-0-135 Retail_Project]$ hadoop fs -ls /tmp/op2
Found 48 items
drwxr-xr-x - ec2-user supergroup 0 2021-02-06 14:05 /tmp/op2/ spark_metadata
-rw-r--r-- 3 ec2-user supergroup 0 2021-02-06 13:54 /tmp/op2/part-00000-1b098f29-bf6a-4517-a308-218fa988110d-c000.json
-rw-r--r-- 3 ec2-user supergroup 0 2021-02-06 14:05 /tmp/op2/part-00000-1c29ede7-8f52-4132-847d-616592ba76c0-c000.json
-rw-r--r-- 3 ec2-user supergroup 0 2021-02-06 13:56 /tmp/op2/part-00000-251b860d-759a-4774-8b77-49a3eddf03f8-c000.json
-rw-r--r-- 3 ec2-user supergroup 0 2021-02-06 13:50 /tmp/op2/part-00000-30fe6670-6caf-47e7-b0d6-0c27a4ff65ab-c000.json
-rw-r--r-- 3 ec2-user supergroup 0 2021-02-06 13:49 /tmp/op2/part-00000-33065964-1f5c-4f86-8da2-107ea126f7cb-c000.json
-rw-r--r-- 3 ec2-user supergroup 0 2021-02-06 14:00 /tmp/op2/part-00000-3a7980c2-31b5-4fb2-ab55-978a7df080b9-c000.json
-rw-r--r-- 3 ec2-user supergroup 0 2021-02-06 14:04 /tmp/op2/part-00000-465bc2fa-d56a-40af-8413-07192a4f01e8-c000.json
-rw-r--r-- 3 ec2-user supergroup 0 2021-02-06 13:47 /tmp/op2/part-00000-4929484e-31e7-4ad4-9c2c-1138837158ff-c000.json
-rw-r--r-- 3 ec2-user supergroup 0 2021-02-06 13:51 /tmp/op2/part-00000-4ef94b73-9427-4768-8ef0-2a4b5ca3b410-c000.json
-rw-r--r-- 3 ec2-user supergroup 0 2021-02-06 13:59 /tmp/op2/part-00000-59926beb-ea62-45f5-a7cc-55c915554313-c000.json
-rw-r--r-- 3 ec2-user supergroup 0 2021-02-06 14:01 /tmp/op2/part-00000-76617d5f-ba49-4bf2-9c15-e326c9b16f06-c000.json
-rw-r--r-- 3 ec2-user supergroup 0 2021-02-06 13:55 /tmp/op2/part-00000-94ab58f7-6135-43ad-abc1-db0c0db0c922-c000.json
-rw-r--r-- 3 ec2-user supergroup 0 2021-02-06 13:53 /tmp/op2/part-00000-97ce1267-ffc2-4706-ace2-025b39173883-c000.json
-rw-r--r-- 3 ec2-user supergroup 0 2021-02-06 13:57 /tmp/op2/part-00000-b3233054-d10c-4671-8524-a45ef774c60d-c000.json
-rw-r--r-- 3 ec2-user supergroup 0 2021-02-06 14:02 /tmp/op2/part-00000-b3b7f6cf-21e6-4cf2-9d3a-17c0a6101b9a-c000.json
-rw-r--r-- 3 ec2-user supergroup 0 2021-02-06 13:58 /tmp/op2/part-00000-c6014d9c-326b-49ca-9789-2e84edfd09f6-c000.json
-rw-r--r-- 3 ec2-user supergroup 169 2021-02-06 14:03 /tmp/op2/part-00000-cbfd15ec-f1c8-470d-b24c-3c17bf3ebac5-c000.json
-rw-r--r-- 3 ec2-user supergroup 170 2021-02-06 13:53 /tmp/op2/part-00004-2dd30b5c-ba2d-4e51-86c0-145f141a8fde-c000.json
-rw-r--r-- 3 ec2-user supergroup 170 2021-02-06 13:51 /tmp/op2/part-00004-5b5fd032-da6f-4e45-8618-324cf0106bd4-c000.json
-rw-r--r-- 3 ec2-user supergroup 170 2021-02-06 13:57 /tmp/op2/part-00031-3ac33e65-94d7-43cd-8f45-550dad744a86-c000.json
-rw-r--r-- 3 ec2-user supergroup 170 2021-02-06 13:54 /tmp/op2/part-00037-a906b081-78f6-41bc-a976-2c38aaa98b6f-c000.json
-rw-r--r-- 3 ec2-user supergroup 157 2021-02-06 13:57 /tmp/op2/part-00040-8aa53e27-015b-49bd-ba0d-dfa36e3585fc-c000.json
-rw-r--r-- 3 ec2-user supergroup 163 2021-02-06 14:04 /tmp/op2/part-00042-f4deb6f3-93fa-4568-af09-13d828da31a5-c000.json
-rw-r--r-- 3 ec2-user supergroup 168 2021-02-06 14:00 /tmp/op2/part-00059-33aa8e08-2be9-48ad-8f90-55ed454d8d1e-c000.json
-rw-r--r-- 3 ec2-user supergroup 169 2021-02-06 13:55 /tmp/op2/part-00060-bb0e26e4-f324-493b-a131-3b95f6c69f94-c000.json
-rw-r--r-- 3 ec2-user supergroup 169 2021-02-06 13:50 /tmp/op2/part-00062-bc4eca21-863a-49ff-9182-faa027d98507-c000.json
-rw-r--r-- 3 ec2-user supergroup 160 2021-02-06 13:58 /tmp/op2/part-00071-a13d7499-a90e-4ba4-b17f-dc4ad538d132-c000.json
-rw-r--r-- 3 ec2-user supergroup 161 2021-02-06 13:59 /tmp/op2/part-00081-d9da4566-9ee9-4708-9fcc-acd3e1f765c0-c000.json
-rw-r--r-- 3 ec2-user supergroup 170 2021-02-06 13:56 /tmp/op2/part-00083-1bf501f0-fd91-4eca-bfec-3d0362abce56-c000.json
-rw-r--r-- 3 ec2-user supergroup 158 2021-02-06 13:55 /tmp/op2/part-00095-b2e92aec-d060-4963-bd3d-3b8079c2e1b2-c000.json
-rw-r--r-- 3 ec2-user supergroup 161 2021-02-06 13:57 /tmp/op2/part-00096-08c031bd-dbd2-423c-b895-232b6ac425a9-c000.json
-rw-r--r-- 3 ec2-user supergroup 168 2021-02-06 13:58 /tmp/op2/part-00098-0da3a27a-5543-443e-a3dd-92f2e6a3e85d-c000.json
-rw-r--r-- 3 ec2-user supergroup 160 2021-02-06 13:54 /tmp/op2/part-00105-b4b1b419-a541-4063-a19d-1f677adb3597-c000.json
-rw-r--r-- 3 ec2-user supergroup 160 2021-02-06 13:50 /tmp/op2/part-00118-c89cf3d4-35ed-44ad-9a0b-5e242d0cfff6a-c000.json
-rw-r--r-- 3 ec2-user supergroup 172 2021-02-06 13:55 /tmp/op2/part-00125-95044a70-95f3-4ec7-9892-c49de76e481b-c000.json
-rw-r--r-- 3 ec2-user supergroup 172 2021-02-06 14:01 /tmp/op2/part-00133-96bc6735-24b2-479e-a1f6-1b67e5d40776-c000.json
-rw-r--r-- 3 ec2-user supergroup 161 2021-02-06 14:05 /tmp/op2/part-00134-a3cf8f73-b38b-4424-84dc-7da0b324af6d-c000.json
-rw-r--r-- 3 ec2-user supergroup 158 2021-02-06 13:59 /tmp/op2/part-00137-13ff80f8-66c8-4377-af0c-a933e958896f-c000.json
-rw-r--r-- 3 ec2-user supergroup 158 2021-02-06 14:00 /tmp/op2/part-00143-57cf565e-bc4e-46f8-87b7-lac33018935d-c000.json
-rw-r--r-- 3 ec2-user supergroup 169 2021-02-06 13:59 /tmp/op2/part-00143-beb04d39-6e0c-45ac-911b-553efa8cd038-c000.json
-rw-r--r-- 3 ec2-user supergroup 159 2021-02-06 13:55 /tmp/op2/part-00154-8700de86-c0cd-4d9b-ae95-323af19bc7dd-c000.json
```

Read record from json file from /tmp/op2.

Hadoop fs -cat /tmp/op2/<json-file-name>

```
ec2-user@ip-10-0-0-135 Retail_Project$ hadoop fs -cat /tmp/op2/part-00154-8700de86-c0cd-4d9b-ae95-323af19bc7dd-c000.json
{"window":{"start":"2021-02-06T13:58:00.000Z","end":"2021-02-06T13:59:00.000Z"},"country":"United Kingdom","OPM":6,"total_sale_volume":"138.29","rate_of_return":"0.00"}
[ec2-user@ip-10-0-0-135 Retail_Project]$
```

3) CP1

Hadoop fs -ls /tmp/cp1

```
[ec2-user@ip-10-0-0-135 Retail_Project]$ hadoop fs -ls /tmp/cp1
Found 5 items
drwxr-xr-x - ec2-user supergroup          0 2021-02-06 14:06 /tmp/cp1/commits
-rw-r--r--  3 ec2-user supergroup        45 2021-02-06 13:47 /tmp/cp1/metadata
drwxr-xr-x - ec2-user supergroup          0 2021-02-06 14:07 /tmp/cp1/offsets
drwxr-xr-x - ec2-user supergroup          0 2021-02-06 13:47 /tmp/cp1/sources
drwxr-xr-x - ec2-user supergroup          0 2021-02-06 13:47 /tmp/cp1/state
[ec2-user@ip-10-0-0-135 Retail_Project]$
```

4) CP2

Hadoop fs -ls /tmp/cp2

```
[ec2-user@ip-10-0-0-135 Retail_Project]$ hadoop fs -ls /tmp/cp2
Found 5 items
drwxr-xr-x - ec2-user supergroup          0 2021-02-06 14:08 /tmp/cp2/commits
-rw-r--r--  3 ec2-user supergroup        45 2021-02-06 13:47 /tmp/cp2/metadata
drwxr-xr-x - ec2-user supergroup          0 2021-02-06 14:09 /tmp/cp2/offsets
drwxr-xr-x - ec2-user supergroup          0 2021-02-06 13:47 /tmp/cp2/sources
drwxr-xr-x - ec2-user supergroup          0 2021-02-06 13:47 /tmp/cp2/state
[ec2-user@ip-10-0-0-135 Retail_Project]$
```