

Code Logic - Retail Data Analysis

In this document, you will describe the code and the overall steps taken to solve the project.

1. Logic

```
Start_spark_session()
Read_raw_data_from_kafka_topic()
Format_raw_data_to_json_data()
Utility_function_with_formula_for_agg_time()
Utility_function_with_formula_for_agg_time_country()
Write_to_console()
```

2. Utility

UTILITY FUNCTIONS

total_item_count: to sum up quantity of items ordered for each invoice
total_cost: get total cost using $\text{quantity} * \text{unit_price}$ for each invoice
is_a_order: return 1 if type is ORDER else 0
is_a_return: return 1 if type is RETURN else 0

3. Streams

order_stream: Input Stream [Raw data]
order_extended_stream – is the stream with the derived columns added to the raw data.
agg_time : Calculated the time-based KPIs with tumbling window of one minute on orders across the globe.
agg_time_country: Calculated the time and country-based KPIs with tumbling window of one minute on orders across the globe.

4. Command

```
spark-submit --packages org.apache.spark:spark-sql-kafka-0-10_2.11:2.4.5 spark-streaming.py
```

5. Output directory

```
[hadoop@ip-172-31-57-4 ~]$ hadoop fs -ls time_countryKPI/
Found 7 items
drwxr-xr-x - hadoop hadoop      0 2022-09-12 18:07 time_countryKPI/_spark_metadata
drwxr-xr-x - hadoop hadoop      0 2022-09-12 18:04 time_countryKPI/cp
-rw-r--r-- 1 hadoop hadoop      0 2022-09-12 18:07 time_countryKPI/part-00000-19ffb1fd-
bfa3-4168-a35a-7cc8b4bde2c8-c000.json
-rw-r--r-- 1 hadoop hadoop      0 2022-09-12 18:04 time_countryKPI/part-00000-21a45802-
90ae-4efb-805c-21dac3d0befe-c000.json
```

```
-rw-r--r-- 1 hadoop hadoop      0 2022-09-12 18:08 time_countryKPI/part-00000-e16bf570-40ca-407d-bedf-d524a3654cac-c000.json
-rw-r--r-- 1 hadoop hadoop      0 2022-09-12 18:06 time_countryKPI/part-00000-f7c7c6d3-bff8-46fe-9600-1e8e4291db0a-c000.json
-rw-r--r-- 1 hadoop hadoop    161 2022-09-12 18:08 time_countryKPI/part-00013-05660512-c8d7-491f-9462-cdd74297d014-c000.json
```

```
[hadoop@ip-172-31-57-4 ~]$ hadoop fs -ls timeKPI/
```

```
Found 7 items
```

```
drwxr-xr-x - hadoop hadoop      0 2022-09-12 18:08 timeKPI/_spark_metadata
drwxr-xr-x - hadoop hadoop      0 2022-09-12 18:04 timeKPI/cp
-rw-r--r-- 1 hadoop hadoop      0 2022-09-12 18:04 timeKPI/part-00000-1ee1af67-26b9-4ae2-891c-6800b977d32b-c000.json
-rw-r--r-- 1 hadoop hadoop      0 2022-09-12 18:07 timeKPI/part-00000-294d6152-b89b-4f48-a64a-dd9ada408a86-c000.json
-rw-r--r-- 1 hadoop hadoop      0 2022-09-12 18:08 timeKPI/part-00000-9d76c9ce-ff0c-4371-9802-6a9d0f7b3b86-c000.json
-rw-r--r-- 1 hadoop hadoop      0 2022-09-12 18:06 timeKPI/part-00000-f550808b-8f24-4465-b95e-15b67f640557-c000.json
-rw-r--r-- 1 hadoop hadoop    177 2022-09-12 18:08 timeKPI/part-00097-e122d6f2-7065-42a4-91db-c2b924570db9-c000.json
```

```
[hadoop@ip-172-31-57-4 ~]$ hadoop fs -cat time_countryKPI/part-00013-05660512-c8d7-491f-9462-cdd74297d014-c000.json
{"start":"2022-09-12T17:59:00.000Z","end":"2022-09-12T18:00:00.000Z","country":"France","OPM":1,"total_volume_of_sales":1.4500000476837158,"rate_of_return":0.0}
```

6. Console Output

```
[hadoop@ip-172-31-57-4 ~]$ spark-submit --packages org.apache.spark:spark-sql-kafka-0-10_2.11:2.4.5 spark-streaming.py
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-a9f766fe-68e7-40ac-b22d-c91d86ee4abe;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
found org.slf4j#slf4j-api;1.7.16 in central
found org.spark-project.spark#unused;1.0.0 in central
downloading https://repo1.maven.org/maven2/org/apache/spark/spark-sql-kafka-0-10_2.11/2.4.5/spark-sql-kafka-0-10_2.11-2.4.5.jar ...
[SUCCESSFUL ] org.apache.spark#spark-sql-kafka-0-10_2.11;2.4.5!spark-sql-kafka-0-10_2.11.jar (43ms)
downloading https://repo1.maven.org/maven2/org/apache/kafka/kafka-clients/2.0.0/kafka-clients-2.0.0.jar ...
[SUCCESSFUL ] org.apache.kafka#kafka-clients;2.0.0!kafka-clients.jar (122ms)
downloading https://repo1.maven.org/maven2/org/spark-project/spark/unused/1.0.0/unused-1.0.0.jar ...
[SUCCESSFUL ] org.spark-project.spark#unused;1.0.0!unused.jar (7ms)
downloading https://repo1.maven.org/maven2/org/lz4/lz4-java/1.4.0/lz4-java-1.4.0.jar ...
[SUCCESSFUL ] org.lz4#lz4-java;1.4.0!lz4-java.jar (32ms)
downloading https://repo1.maven.org/maven2/org/xerial/snappy/snappy-java/1.1.7.3/snappy-java-1.1.7.3.jar ...
[SUCCESSFUL ] org.xerial.snappy#snappy-java;1.1.7.3!snappy-java.jar(bundle) (87ms)
downloading https://repo1.maven.org/maven2/org/slf4j/slf4j-api/1.7.16/slf4j-api-1.7.16.jar ...
[SUCCESSFUL ] org.slf4j#slf4j-api;1.7.16!slf4j-api.jar (5ms)
:: resolution report :: resolve 2530ms :: artifacts dl 311ms
:: modules in use:
org.apache.kafka#kafka-clients;2.0.0 from central in [default]
org.apache.spark#spark-sql-kafka-0-10_2.11;2.4.5 from central in [default]
org.lz4#lz4-java;1.4.0 from central in [default]
org.slf4j#slf4j-api;1.7.16 from central in [default]
org.spark-project.spark#unused;1.0.0 from central in [default]
org.xerial.snappy#snappy-java;1.1.7.3 from central in [default]
-----
|      |      modules      || artifacts |
|  conf  | number| search|dwnlded|evicted|| number|dwnlded|
-----
| default | 6 | 6 | 6 | 0 || 6 | 6 |
-----

:: problems summary ::
::: ERRORS
    SERVER ERROR: Bad Gateway url=https://dl.bintray.com/spark-packages/maven/org/apache/apache/18/apache-18.jar

    SERVER ERROR: Bad Gateway url=https://dl.bintray.com/spark-packages/maven/org/apache/spark/spark-parent_2.11/2.4.5/spark-parent_2.11-2.4.5.jar

```

SERVER ERROR: Bad Gateway url=https://dl.bintray.com/spark-packages/maven/org/apache/spark/spark-sql-kafka-0-10_2.11/2.4.5/spark-sql-kafka-0-10_2.11-2.4.5-javadoc.jar

SERVER ERROR: Bad Gateway url=https://dl.bintray.com/spark-packages/maven/org/slf4j/slf4j-parent/1.7.16/slf4j-parent-1.7.16.jar

SERVER ERROR: Bad Gateway url=https://dl.bintray.com/spark-packages/maven/org/sonatype/oss/oss-parent/9/oss-parent-9.jar

```
:: USE VERBOSE OR DEBUG MESSAGE LEVEL FOR MORE DETAILS
:: retrieving :: org.apache.spark#spark-submit-parent-a9f766fe-68e7-40ac-b22d-c91d86ee4abe
  confs: [default]
  6 artifacts copied, 0 already retrieved (4749kB/29ms)
Traceback (most recent call last):
  File "/home/hadoop/spark-streaming.py", line 7, in <module>
    import findspark
ModuleNotFoundError: No module named 'findspark'
22/09/12 18:03:51 INFO ShutdownHookManager: Shutdown hook called
22/09/12 18:03:51 INFO ShutdownHookManager: Deleting directory /mnt/tmp/spark-b77a4029-a0bb-40df-bfaa-7cae2873a5bb
[hadoop@ip-172-31-57-4 ~]$ vim spark-streaming.py
[hadoop@ip-172-31-57-4 ~]$ spark-submit --packages org.apache.spark:spark-sql-kafka-0-10_2.11:2.4.5 spark-streaming.py
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-b2b9a1d6-f0e1-468c-924b-e3479839dc25;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
  found org.slf4j#slf4j-api;1.7.16 in central
  found org.spark-project.spark#unused;1.0.0 in central
:: resolution report :: resolve 557ms :: artifacts dl 18ms
  :: modules in use:
    org.apache.kafka#kafka-clients;2.0.0 from central in [default]
```

```
org.apache.spark#spark-sql-kafka-0-10_2.11;2.4.5 from central in [default]
org.lz4#lz4-java;1.4.0 from central in [default]
org.slf4j#slf4j-api;1.7.16 from central in [default]
org.spark-project.spark#unused;1.0.0 from central in [default]
org.xerial.snappy#snappy-java;1.1.7.3 from central in [default]
```

		modules				artifacts			
conf	number	search	downloaded	evicted		number	downloaded		
default	6	0	0	0		6	0		

```
:: retrieving :: org.apache.spark#spark-submit-parent-b2b9a1d6-f0e1-468c-924b-e3479839dc25
```

```
confs: [default]
```

```
0 artifacts copied, 6 already retrieved (0kB/13ms)
```

```
22/09/12 18:04:09 INFO SparkContext: Running Spark version 2.4.5-amzn-0
22/09/12 18:04:09 INFO SparkContext: Submitted application: StructuredSocketRead
22/09/12 18:04:09 INFO SecurityManager: Changing view acls to: hadoop
22/09/12 18:04:09 INFO SecurityManager: Changing modify acls to: hadoop
22/09/12 18:04:09 INFO SecurityManager: Changing view acls groups to:
22/09/12 18:04:09 INFO SecurityManager: Changing modify acls groups to:
22/09/12 18:04:09 INFO SecurityManager: SecurityManager: authentication disabled; ui
acls disabled; users with view permissions: Set(hadoop); groups with view permissions:
Set(); users with modify permissions: Set(hadoop); groups with modify permissions:
Set()
22/09/12 18:04:10 INFO Utils: Successfully started service 'sparkDriver' on port 36293.
22/09/12 18:04:10 INFO SparkEnv: Registering MapOutputTracker
22/09/12 18:04:11 INFO SparkEnv: Registering BlockManagerMaster
22/09/12 18:04:11 INFO BlockManagerMasterEndpoint: Using
org.apache.spark.storage.DefaultTopologyMapper for getting topology information
22/09/12 18:04:11 INFO BlockManagerMasterEndpoint: BlockManagerMasterEndpoint
up
22/09/12 18:04:11 INFO DiskBlockManager: Created local directory at
/mnt/tmp/blockmgr-6dba7d3f-5507-43b5-843c-f2544c551761
22/09/12 18:04:11 INFO MemoryStore: MemoryStore started with capacity 1038.8 MB
22/09/12 18:04:11 INFO SparkEnv: Registering OutputCommitCoordinator
22/09/12 18:04:11 INFO Utils: Successfully started service 'SparkUI' on port 4040.
22/09/12 18:04:11 INFO SparkUI: Bound SparkUI to 0.0.0.0, and started at http://ip-172-31-57-4.ec2.internal:4040
22/09/12 18:04:12 INFO Utils: Using initial executors = 50, max of
spark.dynamicAllocation.initialExecutors, spark.dynamicAllocation.minExecutors and
spark.executor.instances
22/09/12 18:04:12 INFO RMPProxy: Connecting to ResourceManager at ip-172-31-57-4.ec2.internal/172.31.57.4:8032
```

22/09/12 18:04:13 INFO Client: Requesting a new application from cluster with 2 NodeManagers

22/09/12 18:04:13 INFO Client: Verifying our application has not requested more than the maximum memory capability of the cluster (6144 MB per container)

22/09/12 18:04:13 INFO Client: Will allocate AM container, with 896 MB memory including 384 MB overhead

22/09/12 18:04:13 INFO Client: Setting up container launch context for our AM

22/09/12 18:04:13 INFO Client: Setting up the launch environment for our AM container

22/09/12 18:04:13 INFO Client: Preparing resources for our AM container

22/09/12 18:04:13 WARN Client: Neither spark.yarn.jars nor spark.yarn.archive is set, falling back to uploading libraries under SPARK_HOME.

22/09/12 18:04:16 INFO Client: Uploading resource file:/mnt/tmp/spark-6b83ccfb-38e4-41f3-82b5-d1003a5c418f/___spark_libs___1030221316877487471.zip -> hdfs://ip-172-31-57-4.ec2.internal:8020/user/hadoop/.sparkStaging/application_1663000676736_0003/___spark_libs___1030221316877487471.zip

22/09/12 18:04:18 INFO Client: Uploading resource file:/home/hadoop/.ivy2/jars/org.apache.spark_spark-sql-kafka-0-10_2.11-2.4.5.jar -> hdfs://ip-172-31-57-4.ec2.internal:8020/user/hadoop/.sparkStaging/application_1663000676736_0003/org.apache.spark_spark-sql-kafka-0-10_2.11-2.4.5.jar

22/09/12 18:04:18 INFO Client: Uploading resource file:/home/hadoop/.ivy2/jars/org.apache.kafka_kafka-clients-2.0.0.jar -> hdfs://ip-172-31-57-4.ec2.internal:8020/user/hadoop/.sparkStaging/application_1663000676736_0003/org.apache.kafka_kafka-clients-2.0.0.jar

22/09/12 18:04:18 INFO Client: Uploading resource file:/home/hadoop/.ivy2/jars/org.spark-project.spark_unused-1.0.0.jar -> hdfs://ip-172-31-57-4.ec2.internal:8020/user/hadoop/.sparkStaging/application_1663000676736_0003/org.spark-project.spark_unused-1.0.0.jar

22/09/12 18:04:18 INFO Client: Uploading resource file:/home/hadoop/.ivy2/jars/org.lz4_lz4-java-1.4.0.jar -> hdfs://ip-172-31-57-4.ec2.internal:8020/user/hadoop/.sparkStaging/application_1663000676736_0003/org.lz4_lz4-java-1.4.0.jar

22/09/12 18:04:18 INFO Client: Uploading resource file:/home/hadoop/.ivy2/jars/org.xerial.snappy_snappy-java-1.1.7.3.jar -> hdfs://ip-172-31-57-4.ec2.internal:8020/user/hadoop/.sparkStaging/application_1663000676736_0003/org.xerial.snappy_snappy-java-1.1.7.3.jar

22/09/12 18:04:18 INFO Client: Uploading resource file:/home/hadoop/.ivy2/jars/org.slf4j_slf4j-api-1.7.16.jar -> hdfs://ip-172-31-57-4.ec2.internal:8020/user/hadoop/.sparkStaging/application_1663000676736_0003/org.slf4j_slf4j-api-1.7.16.jar

22/09/12 18:04:18 INFO Client: Uploading resource
file:/usr/lib/spark/python/lib/pyspark.zip -> hdfs://ip-172-31-57-
4.ec2.internal:8020/user/hadoop/.sparkStaging/application_1663000676736_0003/pysp
ark.zip
22/09/12 18:04:18 INFO Client: Uploading resource file:/usr/lib/spark/python/lib/py4j-
0.10.7-src.zip -> hdfs://ip-172-31-57-
4.ec2.internal:8020/user/hadoop/.sparkStaging/application_1663000676736_0003/py4j-
0.10.7-src.zip
22/09/12 18:04:18 WARN Client: Same path resource
file:///home/hadoop/.ivy2/jars/org.apache.spark_spark-sql-kafka-0-10_2.11-2.4.5.jar
added multiple times to distributed cache.
22/09/12 18:04:18 WARN Client: Same path resource
file:///home/hadoop/.ivy2/jars/org.apache.kafka_kafka-clients-2.0.0.jar added multiple
times to distributed cache.
22/09/12 18:04:18 WARN Client: Same path resource
file:///home/hadoop/.ivy2/jars/org.spark-project.spark_unused-1.0.0.jar added multiple
times to distributed cache.
22/09/12 18:04:18 WARN Client: Same path resource
file:///home/hadoop/.ivy2/jars/org.lz4_lz4-java-1.4.0.jar added multiple times to
distributed cache.
22/09/12 18:04:18 WARN Client: Same path resource
file:///home/hadoop/.ivy2/jars/org.xerial.snappy_snappy-java-1.1.7.3.jar added multiple
times to distributed cache.
22/09/12 18:04:18 WARN Client: Same path resource
file:///home/hadoop/.ivy2/jars/org.slf4j_slf4j-api-1.7.16.jar added multiple times to
distributed cache.
22/09/12 18:04:18 INFO Client: Uploading resource file:/mnt/tmp/spark-6b83ccfb-38e4-
41f3-82b5-d1003a5c418f/__spark_conf__5148411904578300620.zip -> hdfs://ip-172-
31-57-
4.ec2.internal:8020/user/hadoop/.sparkStaging/application_1663000676736_0003/__sp
ark_conf__.zip
22/09/12 18:04:18 INFO SecurityManager: Changing view acls to: hadoop
22/09/12 18:04:18 INFO SecurityManager: Changing modify acls to: hadoop
22/09/12 18:04:18 INFO SecurityManager: Changing view acls groups to:
22/09/12 18:04:18 INFO SecurityManager: Changing modify acls groups to:
22/09/12 18:04:18 INFO SecurityManager: SecurityManager: authentication disabled; ui
acls disabled; users with view permissions: Set(hadoop); groups with view permissions:
Set(); users with modify permissions: Set(hadoop); groups with modify permissions:
Set()
22/09/12 18:04:20 INFO Client: Submitting application
application_1663000676736_0003 to ResourceManager
22/09/12 18:04:20 INFO YarnClientImpl: Submitted application
application_1663000676736_0003

22/09/12 18:04:20 INFO SchedulerExtensionServices: Starting Yarn extension services with app application_1663000676736_0003 and attemptId None

22/09/12 18:04:22 INFO Client: Application report for application_1663000676736_0003 (state: ACCEPTED)

22/09/12 18:04:22 INFO Client:

client token: N/A

diagnostics: AM container is launched, waiting for AM container to Register with

RM

ApplicationMaster host: N/A

ApplicationMaster RPC port: -1

queue: default

start time: 1663005860828

final status: UNDEFINED

tracking URL: http://ip-172-31-57-

4.ec2.internal:20888/proxy/application_1663000676736_0003/

user: hadoop

22/09/12 18:04:23 INFO Client: Application report for application_1663000676736_0003 (state: ACCEPTED)

22/09/12 18:04:24 INFO Client: Application report for application_1663000676736_0003 (state: ACCEPTED)

22/09/12 18:04:25 INFO Client: Application report for application_1663000676736_0003 (state: ACCEPTED)

22/09/12 18:04:26 INFO Client: Application report for application_1663000676736_0003 (state: ACCEPTED)

22/09/12 18:04:27 INFO Client: Application report for application_1663000676736_0003 (state: ACCEPTED)

22/09/12 18:04:28 INFO YarnClientSchedulerBackend: Add WebUI Filter.

org.apache.hadoop.yarn.server.webproxy.amfilter.AmIpFilter, Map(PROXY_HOSTS ->

ip-172-31-57-4.ec2.internal, PROXY_URI_BASES -> http://ip-172-31-57-

4.ec2.internal:20888/proxy/application_1663000676736_0003),

/proxy/application_1663000676736_0003

22/09/12 18:04:28 INFO Client: Application report for application_1663000676736_0003 (state: RUNNING)

22/09/12 18:04:28 INFO Client:

client token: N/A

diagnostics: N/A

ApplicationMaster host: 172.31.50.9

ApplicationMaster RPC port: -1

queue: default

start time: 1663005860828

final status: UNDEFINED

tracking URL: http://ip-172-31-57-

4.ec2.internal:20888/proxy/application_1663000676736_0003/

user: hadoop

22/09/12 18:04:28 INFO YarnClientSchedulerBackend: Application application_1663000676736_0003 has started running.
22/09/12 18:04:28 INFO Utils: Successfully started service 'org.apache.spark.network.netty.NettyBlockTransferService' on port 42661.
22/09/12 18:04:28 INFO NettyBlockTransferService: Server created on ip-172-31-57-4.ec2.internal:42661
22/09/12 18:04:28 INFO BlockManager: Using org.apache.spark.storage.RandomBlockReplicationPolicy for block replication policy
22/09/12 18:04:28 INFO BlockManagerMaster: Registering BlockManager BlockManagerId(driver, ip-172-31-57-4.ec2.internal, 42661, None)
22/09/12 18:04:28 INFO BlockManagerMasterEndpoint: Registering block manager ip-172-31-57-4.ec2.internal:42661 with 1038.8 MB RAM, BlockManagerId(driver, ip-172-31-57-4.ec2.internal, 42661, None)
22/09/12 18:04:28 INFO BlockManagerMaster: Registered BlockManager BlockManagerId(driver, ip-172-31-57-4.ec2.internal, 42661, None)
22/09/12 18:04:28 INFO BlockManager: external shuffle service port = 7337
22/09/12 18:04:28 INFO BlockManager: Initialized BlockManager: BlockManagerId(driver, ip-172-31-57-4.ec2.internal, 42661, None)
22/09/12 18:04:28 INFO YarnSchedulerBackend\$YarnSchedulerEndpoint: ApplicationMaster registered as NettyRpcEndpointRef(spark-client://YarnAM)
22/09/12 18:04:28 INFO JettyUtils: Adding filter org.apache.hadoop.yarn.server.webproxy.amfilter.AmIpFilter to /metrics/json.
22/09/12 18:04:29 INFO EventLoggingListener: Logging events to hdfs:/var/log/spark/apps/application_1663000676736_0003
22/09/12 18:04:29 INFO Utils: Using initial executors = 50, max of spark.dynamicAllocation.initialExecutors, spark.dynamicAllocation.minExecutors and spark.executor.instances
22/09/12 18:04:30 INFO YarnClientSchedulerBackend: SchedulerBackend is ready for scheduling beginning after reached minRegisteredResourcesRatio: 0.0
22/09/12 18:04:30 INFO SharedState: loading hive config file: file:/etc/spark/conf.dist/hive-site.xml
22/09/12 18:04:30 INFO SharedState: Setting hive.metastore.warehouse.dir ('null') to the value of spark.sql.warehouse.dir ('hdfs:///user/spark/warehouse').
22/09/12 18:04:30 INFO SharedState: Warehouse path is 'hdfs:///user/spark/warehouse'.
22/09/12 18:04:30 INFO JettyUtils: Adding filter org.apache.hadoop.yarn.server.webproxy.amfilter.AmIpFilter to /SQL.
22/09/12 18:04:30 INFO JettyUtils: Adding filter org.apache.hadoop.yarn.server.webproxy.amfilter.AmIpFilter to /SQL/json.
22/09/12 18:04:30 INFO JettyUtils: Adding filter org.apache.hadoop.yarn.server.webproxy.amfilter.AmIpFilter to /SQL/execution.
22/09/12 18:04:30 INFO JettyUtils: Adding filter org.apache.hadoop.yarn.server.webproxy.amfilter.AmIpFilter to /SQL/execution/json.
22/09/12 18:04:30 INFO JettyUtils: Adding filter org.apache.hadoop.yarn.server.webproxy.amfilter.AmIpFilter to /static/sql.

22/09/12 18:04:32 INFO StateStoreCoordinatorRef: Registered StateStoreCoordinator endpoint

Batch: 0

invoice_no	country	timestamp	type	total_items	total_cost	is_order	is_return

Batch: 1

invoice_no	country	timestamp	type	total_items	total_cost	is_order	is_return
154132551175232	United Kingdom	2022-09-12 17:59:34	ORDER	13	53.29	1	0
154132551175233	United Kingdom	2022-09-12 17:59:42	ORDER	46	56.74	1	0
154132551175234	United Kingdom	2022-09-12 17:59:44	ORDER	44	80.25	1	0
154132551175235	United Kingdom	2022-09-12 17:59:44	ORDER	29	90.6	1	0
154132551175236	France	2022-09-12 17:59:48	ORDER	1	1.45	1	0
154132551175237	United Kingdom	2022-09-12 17:59:56	ORDER	39	103.9	1	0
154132551175238	United Kingdom	2022-09-12 17:59:56	ORDER	1	2.95	1	0
154132551175239	Switzerland	2022-09-12 17:59:58	ORDER	4	16.5	1	0
154132551175240	United Kingdom	2022-09-12 18:00:05	ORDER	59	98.799995	1	0
154132551175241	United Kingdom	2022-09-12 18:00:11	ORDER	2	9.88	1	0
154132551175242	United Kingdom	2022-09-12 18:00:27	ORDER	64	95.4	1	0

22/09/12 18:06:04 ERROR TransportResponseHandler: Still have 1 requests outstanding when connection from /172.31.61.20:59250 is closed

Batch: 2

invoice_no	country	timestamp	type	total_items	total_cost	is_order	is_return
154132551175243	United Kingdom	2022-09-12 18:00:45	RETURN	2	-4.5	0	1
154132551175244	United Kingdom	2022-09-12 18:00:47	RETURN	46	-139.92	0	1
154132551175245	United Kingdom	2022-09-12 18:00:50	ORDER	78	122.46	1	0
154132551175246	United Kingdom	2022-09-12 18:00:55	RETURN	3	-	14.849999	0
154132551175247	France	2022-09-12 18:00:56	ORDER	50	37.37	1	0
154132551175248	United Kingdom	2022-09-12 18:01:06	ORDER	24	20.64	1	0
154132551175249	United Kingdom	2022-09-12 18:01:12	RETURN	10	-16.46	0	1
154132551175250	United Kingdom	2022-09-12 18:01:15	ORDER	11	33.49	1	0
154132551175251	United Kingdom	2022-09-12 18:01:25	ORDER	25	11.769999	1	0
154132551175252	United Kingdom	2022-09-12 18:01:26	ORDER	88	187.48	1	0
154132551175253	United Kingdom	2022-09-12 18:01:27	ORDER	3	13.65	1	0
154132551175254	United Kingdom	2022-09-12 18:01:27	ORDER	12	29.52	1	0

Batch: 3

invoice_no	country	timestamp	type	total_items	total_cost	is_order	is_return
154132551175255	United Kingdom	2022-09-12 18:01:40	ORDER	30	84.88	1	0
154132551175256	United Kingdom	2022-09-12 18:01:49	ORDER	30	37.94	1	0

154132551175257 United Kingdom 2022-09-12 18:01:55 ORDER 21	69.75	1
0		
154132551175258 United Kingdom 2022-09-12 18:02:10 ORDER 28	39.53	1
0		
154132551175259 United Kingdom 2022-09-12 18:02:14 ORDER 3	5.0299997	1
0		
154132551175260 United Kingdom 2022-09-12 18:02:19 ORDER 83	187.31	1
0		
154132551175261 United Kingdom 2022-09-12 18:02:20 ORDER 14	23.91	1
0		
154132551175262 United Kingdom 2022-09-12 18:02:22 ORDER 39	38.75	1
0		
154132551175263 United Kingdom 2022-09-12 18:02:26 ORDER 72	151.2	1
0		
154132551175264 United Kingdom 2022-09-12 18:02:29 ORDER 81	203.87	1
0		
+-----+-----+-----+-----+-----+-----+		