

Code Logic - Retail Data Analysis

PySpark Code Logic

Step 1: Importing the required modules - *System dependencies for CDH* and declaring PySpark environment variables

importing the modules - System dependencies for CDH

```
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
```

Step 2: Creating Initial Spark Session

Initializing Spark Session

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

Step 3: Reading the Input from Kafka

Read Input from kafka

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

Step 4: Defining the schema for incoming data JSON file data

Defining the Schema

```
jsonSchema = StructType() \
    .add("invoice_no", LongType()) \
    .add("country", StringType()) \
    .add("timestamp", TimestampType()) \
    .add("type", StringType()) \
    .add("items", ArrayType(StructType([
        StructField("SKU", StringType()),
        StructField("title", StringType()),
        StructField("unit_price", DoubleType()),
        StructField("quantity", IntegerType())
    ])))
```

Step 5: Creating dataframe from incoming data

Creating dataframe from input data after applying the schema

```
orderStream = orderRaw.select(from_json(col("value").cast("string"),
jsonSchema).alias("data")).select("data.*")
```

Step 6: Defining UDF functions (total items, total cost, is order, is return), conversion to UDF types and calculating columns

UDF for calculating total_items

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

UDF for calculating order type

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

UDF for calculating return type

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

UDF for calculating total_cost

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

Converting to UDF's with the utility functions

```
isorder = udf(is_order, IntegerType())  
isreturn = udf(is_return, IntegerType())  
totalcount = udf(items_TotalCount, IntegerType())  
totalcost = udf(TotalCostSum, DoubleType())
```

Calculating columns(total_cost, total_items, is_order, is_return)

```
order_stream = orderStream \  
    .withColumn("total_cost", totalcost(orderStream.items, orderStream.type)) \  
    .withColumn("total_items", totalcount(orderStream.items)) \  
    .withColumn("is_order", isorder(orderStream.type)) \  
    .withColumn("is_return", isreturn(orderStream.type))
```

Step 7: Writing intermediate dataset to console with 1 Minute Interval

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

Step 8: Calculating time-based KPIs using with Watermark and groupBy

Calculating time based KPIs

```
timeBasedKPIs = order_stream \  
    .withWatermark("timestamp", "1 minute") \  
    .groupBy(window("timestamp", "1 minute", "1 minute")) \  
    .agg(count("invoice_no").alias("OPM"),  
        sum("total_cost").alias("total_sales_volume"),  
        avg("total_cost").alias("average_transaction_size"),  
        avg("is_return").alias("rate_of_return")) \  
    .select("window", "OPM", "total_sales_volume", "average_transaction_size",  
    "rate_of_return")
```

write stream for time based KPIs

```
timeBasedKPIsOutput = timeBasedKPIs \
    .writeStream \
    .outputMode("append") \
    .format("json") \
    .option("format","append") \
    .option("truncate", "false") \
    .option("path", "time-wise-kpi") \
    .option("checkpointLocation", "time-kpi") \
    .trigger(processingTime="1 minute") \
    .start()
```

Step 9: Calculating time-based and country-based KPIs using watermark and groupBy

Calculating time-based and country-based KPIs

```
timeAndCountryBasedKPIs = order_stream \
    .withWatermark("timestamp", "1 minute") \
    .groupBy(window("timestamp", "1 minute", "1 minute"), "country") \
    .agg(count("invoice_no").alias("OPM"),
        sum("total_cost").alias("total_sales_volume"),
        avg("is_return").alias("rate_of_return")) \
    .select("window", "country", "OPM", "total_sales_volume", "rate_of_return")
```

write stream for time and country based KPIs

```
timeAndCountryBasedKPIsOutput = timeAndCountryBasedKPIs \
    .writeStream \
    .outputMode("Append") \
    .format("json") \
    .option("format","append") \
    .option("truncate", "false") \
    .option("path", "time-country-wise-kpi") \
    .option("checkpointLocation", "time-country-kpi") \
```

```
.trigger(processingTime="1 minute") \  
.start()
```

Step 10: Waiting for the termination of stream infinitely

Waiting infinitely to read the data

```
timeAndCountryBasedKPIsOutput.awaitTermination()
```

Console Commands and Analysis

Spark Submit Command and too Console-output file

```
Spark2-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 > Console-output
```

Mkdir command to make directory

```
mkdir time-wise-kpi  
mkdir time-country-wise-kpi
```

command to check json files generated

```
hadoop fs -ls
```

```
hadoop fs -ls time-wise-kpi  
hadoop fs -ls time-country-wise-kpi
```

```
hadoop fs -cat time-wise-kpi/part-00198-21224b93-1ac7-4748-806b-d39e02d0b82a-c000/  
hadoop fs -cat time-country-wise-kpi/part-00185-ffa2a99d-0618-41c2-abab-cb73865996b7-  
c000/
```

Transfer of data from HDFS to Local using

- `hadoop fs -get /user/hadoop/time-wise-kpi ./time-wise-kpi`
- `hadoop fs -get /user/hadoop/time-country-wise-kpi ./time-country-wise-kpi`
- WinSCP used to copy the files to local machine

Output Screens Summarized Console Output

```
hadoop@ip-172-31-6-228:~
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|
|-----+-----+-----+-----+-----+-----+
|154132549897430|United Kingdom|2022-06-15 16:01:00|45.54|24|1|0|
|154132549897431|United Kingdom|2022-06-15 16:01:01|3.9000000000000004|4|1|0|
|154132549897432|United Kingdom|2022-06-15 16:01:06|-16.26|18|0|1|
|154132549897433|France|2022-06-15 16:01:08|528.64|128|1|0|
|154132549897434|United Kingdom|2022-06-15 16:01:13|8.32|4|1|0|
|154132549897435|United Kingdom|2022-06-15 16:01:18|54.13|20|1|0|
|154132549897436|United Kingdom|2022-06-15 16:01:31|56.350000000000001|19|1|0|
|154132549897437|United Kingdom|2022-06-15 16:01:32|42.55|21|1|0|
|154132549897438|United Kingdom|2022-06-15 16:01:35|196.40999999999997|53|1|0|
|154132549897439|United Kingdom|2022-06-15 16:01:42|-278.40000000000003|53|0|1|
|154132549897440|United Kingdom|2022-06-15 16:01:42|4.95|1|1|0|
|154132549897441|United Kingdom|2022-06-15 16:01:54|214.570000000000002|177|1|0|
|-----+-----+-----+-----+-----+-----+
Batch: 2
-----+-----+-----+-----+-----+-----+
|invoice_no|country|timestamp|total_cost|total_items|is_order|is_return|
|-----+-----+-----+-----+-----+-----+
|154132549897442|United Kingdom|2022-06-15 16:01:57|89.110000000000001|44|1|0|
|154132549897443|United Kingdom|2022-06-15 16:02:04|44.48|10|1|0|
|154132549897444|United Kingdom|2022-06-15 16:02:05|83.67|43|1|0|
|154132549897445|United Kingdom|2022-06-15 16:02:21|258.33999999999999|108|1|0|
|154132549897446|United Kingdom|2022-06-15 16:02:47|31.9|16|1|0|
|154132549897447|United Kingdom|2022-06-15 16:02:51|165.06999999999996|122|1|0|
|-----+-----+-----+-----+-----+-----+
Batch: 3
-----+-----+-----+-----+-----+-----+
|invoice_no|country|timestamp|total_cost|total_items|is_order|is_return|
|-----+-----+-----+-----+-----+-----+
|154132549897448|United Kingdom|2022-06-15 16:02:52|13.629999999999999|4|1|0|
|154132549897449|United Kingdom|2022-06-15 16:03:09|16.580000000000002|13|1|0|
|154132549897450|Netherlands|2022-06-15 16:03:22|19.349999999999998|14|1|0|
|154132549897451|United Kingdom|2022-06-15 16:03:22|38.410000000000004|15|1|0|
|154132549897452|United Kingdom|2022-06-15 16:03:24|24.55|11|1|0|
|154132549897453|United Kingdom|2022-06-15 16:03:33|9.84|4|1|0|
|154132549897454|United Kingdom|2022-06-15 16:03:34|46.68|42|1|0|
|154132549897455|United Kingdom|2022-06-15 16:03:35|6.76|4|1|0|
|154132549897456|United Kingdom|2022-06-15 16:03:40|24.78|15|1|0|
|154132549897457|United Kingdom|2022-06-15 16:03:45|70.63|19|1|0|
|154132549897458|United Kingdom|2022-06-15 16:03:46|4.97|3|1|0|
|154132549897459|United Kingdom|2022-06-15 16:03:46|-30.5|10|0|1|
|154132549897460|United Kingdom|2022-06-15 16:03:46|148.82999999999998|56|1|0|
|154132549897461|United Kingdom|2022-06-15 16:03:50|30.0|12|1|0|
|-----+-----+-----+-----+-----+-----+
Batch: 4
```

```

Batch: 15
+-----+-----+-----+-----+-----+-----+
|invoice_no|country|timestamp|total_cost|total_items|is_order|is_return|
+-----+-----+-----+-----+-----+-----+
|154132549897584|United Kingdom|2022-06-15 16:14:42|32.480000000000004|6|1|0|
|154132549897585|United Kingdom|2022-06-15 16:14:49|29.52|12|1|0|
|154132549897586|United Kingdom|2022-06-15 16:14:51|4.92|2|1|0|
|154132549897587|United Kingdom|2022-06-15 16:15:02|18.2|2|1|0|
|154132549897588|United Kingdom|2022-06-15 16:15:04|165.18|146|1|0|
|154132549897589|United Kingdom|2022-06-15 16:15:07|102.75|82|1|0|
|154132549897590|United Kingdom|2022-06-15 16:15:09|55.599999999999994|34|1|0|
|154132549897591|United Kingdom|2022-06-15 16:15:13|60.690000000000005|26|1|0|
|154132549897592|United Kingdom|2022-06-15 16:15:18|33.0|20|1|0|
|154132549897593|United Kingdom|2022-06-15 16:15:18|105.28|38|1|0|
|154132549897594|United Kingdom|2022-06-15 16:15:23|10.0|7|1|0|
|154132549897595|United Kingdom|2022-06-15 16:15:26|20.16|6|1|0|
|154132549897596|EIRE|2022-06-15 16:15:27|15.0|12|1|0|
|154132549897597|United Kingdom|2022-06-15 16:15:30|2.89|1|1|0|
+-----+-----+-----+-----+-----+-----+

```

```

^CTraceback (most recent call last):
  File "/home/hadoop/spark-streaming.py", line 161, in <module>
    timeAndCountryBasedKPIsOutput.awaitTermination()
  File "/usr/lib/spark/python/lib/pyspark.zip/pyspark/sql/streaming.py", line 103, in awaitTermination
  File "/usr/lib/spark/python/lib/py4j-0.10.7-src.zip/py4j/java_gateway.py", line 1255, in __call__
  File "/usr/lib/spark/python/lib/py4j-0.10.7-src.zip/py4j/java_gateway.py", line 985, in send_command
  File "/usr/lib/spark/python/lib/py4j-0.10.7-src.zip/py4j/java_gateway.py", line 1152, in send_command
  File "/usr/lib64/python3.7/socket.py", line 589, in readinto
    return self._sock.recv_into(b)
  File "/usr/lib/spark/python/lib/pyspark.zip/pyspark/context.py", line 278, in signal_handler
KeyboardInterrupt
22/06/15 16:19:02 ERROR FileFormatWriter: Aborting job 3351862a-e8d6-4fea-b8b8-06b89f0f5d85.
org.apache.spark.SparkException: Job 63 cancelled as part of cancellation of all jobs
    at org.apache.spark.scheduler.DAGScheduler.org$apache$spark$scheduler$DAGScheduler$$failJobAndIndependentStages(DAGScheduler.scala:2043)
    at org.apache.spark.scheduler.DAGScheduler.handleJobCancellation(DAGScheduler.scala:1978)
    at org.apache.spark.scheduler.DAGScheduler$$anonfun$doCancelAllJobs$1.apply$mcV$fp(DAGScheduler.scala:871)
    at org.apache.spark.scheduler.DAGScheduler$$anonfun$doCancelAllJobs$1.apply(DAGScheduler.scala:871)
    at org.apache.spark.scheduler.DAGScheduler$$anonfun$doCancelAllJobs$1.apply(DAGScheduler.scala:871)
    at scala.collection.mutable.HashSet.foreach(HashSet.scala:78)
    at org.apache.spark.scheduler.DAGScheduler.doCancelAllJobs(DAGScheduler.scala:871)
    at org.apache.spark.scheduler.DAGSchedulerEventProcessLoop.doOnReceive(DAGScheduler.scala:2236)

```

```

[hadoop@ip-172-31-6-228 ~]$ ls
spark-streaming.py
[hadoop@ip-172-31-6-228 ~]$
[hadoop@ip-172-31-6-228 ~]$ mkdir time-wise-kpi
[hadoop@ip-172-31-6-228 ~]$ mkdir time-country-wise-kpi
[hadoop@ip-172-31-6-228 ~]$
[hadoop@ip-172-31-6-228 ~]$
[hadoop@ip-172-31-6-228 ~]$
[hadoop@ip-172-31-6-228 ~]$
[hadoop@ip-172-31-6-228 ~]$ 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-172-31-6-228 ~]$ hadoop fs -ls
Found 5 items
drwxr-xr-x - hadoop hadoop 0 2022-06-15 16:19 .sparkStaging
drwxr-xr-x - hadoop hadoop 0 2022-06-15 16:04 time-country-kpi
drwxr-xr-x - hadoop hadoop 0 2022-06-15 16:18 time-country-wise-kpi
drwxr-xr-x - hadoop hadoop 0 2022-06-15 16:04 time-kpi
drwxr-xr-x - hadoop hadoop 0 2022-06-15 16:19 time-wise-kpi
[hadoop@ip-172-31-6-228 ~]$

```



```
[hadoop@ip-172-31-6-228 ~]$ hadoop fs -ls
Found 5 items
drwxr-xr-x - hadoop hadoop 0 2022-06-15 16:19 .sparkStaging
drwxr-xr-x - hadoop hadoop 0 2022-06-15 16:04 time-country-kpi
drwxr-xr-x - hadoop hadoop 0 2022-06-15 16:18 time-country-wise-kpi
drwxr-xr-x - hadoop hadoop 0 2022-06-15 16:04 time-kpi
drwxr-xr-x - hadoop hadoop 0 2022-06-15 16:19 time-wise-kpi
[hadoop@ip-172-31-6-228 ~]$ hadoop fs -ls time-wise-kpi
Found 28 items
drwxr-xr-x - hadoop hadoop 0 2022-06-15 16:18 time-wise-kpi/_spark_metadata
-rw-r--r-- 1 hadoop hadoop 0 2022-06-15 16:14 time-wise-kpi/part-00000-021e69bd-3cea-4d04-b150-3e0555fceb7b-c000.json
-rw-r--r-- 1 hadoop hadoop 0 2022-06-15 16:15 time-wise-kpi/part-00000-041ea273-20ce-458d-b1d1-8cdaa8de5684-c000.json
-rw-r--r-- 1 hadoop hadoop 0 2022-06-15 16:08 time-wise-kpi/part-00000-044b7eed-e577-40a7-8ce0-049ad71f3231-c000.json
-rw-r--r-- 1 hadoop hadoop 0 2022-06-15 16:05 time-wise-kpi/part-00000-18a9fa97-1fcb-45dc-9758-2825d16f62bc0-c000.json
-rw-r--r-- 1 hadoop hadoop 0 2022-06-15 16:07 time-wise-kpi/part-00000-1ff10c56-1736-438e-bc2a-dc033d55c4d-c000.json
-rw-r--r-- 1 hadoop hadoop 0 2022-06-15 16:11 time-wise-kpi/part-00000-461e169e-a761-410c-aa8b-7c39ba3219f9-c000.json
-rw-r--r-- 1 hadoop hadoop 0 2022-06-15 16:19 time-wise-kpi/part-00000-4c57f2bd-1d14-4682-bb54-5282456c30e7-c000.json
-rw-r--r-- 1 hadoop hadoop 0 2022-06-15 16:06 time-wise-kpi/part-00000-707c36c6-064d-4b14-ba70-2c5097151913-c000.json
-rw-r--r-- 1 hadoop hadoop 0 2022-06-15 16:04 time-wise-kpi/part-00000-76e4dbcc-5dc6-4592-8795-2215d69b017c-c000.json
-rw-r--r-- 1 hadoop hadoop 0 2022-06-15 16:10 time-wise-kpi/part-00000-7e61af6e-a000-4e79-4a6a-352965e1f747-c000.json
-rw-r--r-- 1 hadoop hadoop 0 2022-06-15 16:09 time-wise-kpi/part-00000-84c81f63-4f94-4980-84fd-87eae935862e-c000.json
-rw-r--r-- 1 hadoop hadoop 0 2022-06-15 16:17 time-wise-kpi/part-00000-9af1306e-cfd8-4fae-a807-3a4cbb818185-c000.json
-rw-r--r-- 1 hadoop hadoop 0 2022-06-15 16:12 time-wise-kpi/part-00000-a5e7f1e5-1c93-402f-961f-87adf97041da-c000.json
-rw-r--r-- 1 hadoop hadoop 0 2022-06-15 16:13 time-wise-kpi/part-00000-bc7b91d7-20af-499e-aabf-2eb8cb39d570-c000.json
-rw-r--r-- 1 hadoop hadoop 0 2022-06-15 16:18 time-wise-kpi/part-00000-d3334c4d-4c11-4bef-a0ed-b000fac8d6a3-c000.json
-rw-r--r-- 1 hadoop hadoop 0 2022-06-15 16:16 time-wise-kpi/part-00000-daf43a5-acea-4340-9d0d-8698c3a40771-c000.json
-rw-r--r-- 1 hadoop hadoop 212 2022-06-15 16:12 time-wise-kpi/part-00004-0341bcd9-6bf2-4d98-840f-f277aedd81cf-c000.json
-rw-r--r-- 1 hadoop hadoop 194 2022-06-15 16:09 time-wise-kpi/part-00012-0f16fdd4-2b77-435c-a9e6-e051316d707c-c000.json
-rw-r--r-- 1 hadoop hadoop 201 2022-06-15 16:10 time-wise-kpi/part-00054-48e3dd59-f404-463c-b48d-bb9b3cf0cdcl-c000.json
-rw-r--r-- 1 hadoop hadoop 196 2022-06-15 16:17 time-wise-kpi/part-00062-7d911238-3e40-4d11-843f-3f8d7256e948-c000.json
-rw-r--r-- 1 hadoop hadoop 195 2022-06-15 16:13 time-wise-kpi/part-00087-80fe8e25-b743-408b-811e-0262051dec0f-c000.json
-rw-r--r-- 1 hadoop hadoop 186 2022-06-15 16:18 time-wise-kpi/part-00088-b73849a3-c467-42f3-bb72-75f0f0b92d43-c000.json
-rw-r--r-- 1 hadoop hadoop 196 2022-06-15 16:14 time-wise-kpi/part-00099-0b1a7e0d-9777-4b29-b7d5-e76266e967b1-c000.json
-rw-r--r-- 1 hadoop hadoop 184 2022-06-15 16:15 time-wise-kpi/part-00106-5b2d8b71-4acd-44da-8c6b-fa4892063e23-c000.json
-rw-r--r-- 1 hadoop hadoop 200 2022-06-15 16:11 time-wise-kpi/part-00133-6a735a5d-2429-4e1a-858c-e433640e3d76-c000.json
-rw-r--r-- 1 hadoop hadoop 200 2022-06-15 16:08 time-wise-kpi/part-00154-203a6b62-7fee-4c10-8058-8d5b1be82a84-c000.json
-rw-r--r-- 1 hadoop hadoop 184 2022-06-15 16:16 time-wise-kpi/part-00198-21224b93-lac7-4748-806b-d39e02d0b82a-c000.json
[hadoop@ip-172-31-6-228 ~]$
```

```
[hadoop@ip-172-31-6-228 ~]$ hadoop fs -ls time-country-wise-kpi
Found 38 items
drwxr-xr-x - hadoop hadoop 0 2022-06-15 16:18 time-country-wise-kpi/_spark_metadata
-rw-r--r-- 1 hadoop hadoop 0 2022-06-15 16:14 time-country-wise-kpi/part-00000-0196de06-0879-4821-be0e-081ec5c256f8-c000.json
-rw-r--r-- 1 hadoop hadoop 0 2022-06-15 16:15 time-country-wise-kpi/part-00000-25284aab-6393-40a2-bf36-a189be1bf1cd-c000.json
-rw-r--r-- 1 hadoop hadoop 0 2022-06-15 16:17 time-country-wise-kpi/part-00000-32296765-2dd2-42bb-91ee-a13bdf1e8327-c000.json
-rw-r--r-- 1 hadoop hadoop 0 2022-06-15 16:04 time-country-wise-kpi/part-00000-3519f0a1-5b0d-4cd2-9a10-119a5a0e23ed-c000.json
-rw-r--r-- 1 hadoop hadoop 0 2022-06-15 16:07 time-country-wise-kpi/part-00000-45c70e01-c04e-46ed-872d-c713e0e7f6ld-c000.json
-rw-r--r-- 1 hadoop hadoop 0 2022-06-15 16:18 time-country-wise-kpi/part-00000-7c4aelc2-8500-4673-bf5f-1c906bad6135-c000.json
-rw-r--r-- 1 hadoop hadoop 0 2022-06-15 16:05 time-country-wise-kpi/part-00000-831b5819-7a4e-4212-ba35-2e78ac0d0421-c000.json
-rw-r--r-- 1 hadoop hadoop 0 2022-06-15 16:13 time-country-wise-kpi/part-00000-83cal2d0-59d2-4348-8885-0319428874bc-c000.json
-rw-r--r-- 1 hadoop hadoop 0 2022-06-15 16:10 time-country-wise-kpi/part-00000-8b88acea-afbe-4441-8b2e-0ef5fcff16af-c000.json
-rw-r--r-- 1 hadoop hadoop 0 2022-06-15 16:16 time-country-wise-kpi/part-00000-a237f2cb-7521-4a8c-bca3-615a64b3cb3a-c000.json
-rw-r--r-- 1 hadoop hadoop 0 2022-06-15 16:08 time-country-wise-kpi/part-00000-a724ac59-bc11-4b95-90f5-33f17d86bf62-c000.json
-rw-r--r-- 1 hadoop hadoop 0 2022-06-15 16:06 time-country-wise-kpi/part-00000-de17eb37-5f53-4098-9379-b39f45c39127-c000.json
-rw-r--r-- 1 hadoop hadoop 0 2022-06-15 16:11 time-country-wise-kpi/part-00000-df236787-21b6-4aa8-9647-022f669e29ed-c000.json
-rw-r--r-- 1 hadoop hadoop 0 2022-06-15 16:09 time-country-wise-kpi/part-00000-e6353d2d-5c90-483d-ab9e-8c01319f12d7-c000.json
-rw-r--r-- 1 hadoop hadoop 0 2022-06-15 16:12 time-country-wise-kpi/part-00000-f57ea320-21d3-47e9-adbb-e76562979eea-c000.json
-rw-r--r-- 1 hadoop hadoop 156 2022-06-15 16:17 time-country-wise-kpi/part-00002-3f65b776-731d-4f39-b3bd-5fd1c2841a61-c000.json
-rw-r--r-- 1 hadoop hadoop 176 2022-06-15 16:09 time-country-wise-kpi/part-00012-c344fba-31b1-414a-9b7b-99173af7b5d1-c000.json
-rw-r--r-- 1 hadoop hadoop 155 2022-06-15 16:12 time-country-wise-kpi/part-00019-befc29a8-8460-4b5c-87b9-c7db233ada10-c000.json
-rw-r--r-- 1 hadoop hadoop 169 2022-06-15 16:12 time-country-wise-kpi/part-00022-4c57b3e6-144b-42ff-aeb6-68485b6835f0-c000.json
-rw-r--r-- 1 hadoop hadoop 155 2022-06-15 16:11 time-country-wise-kpi/part-00034-a75939a0-8103-4b5b-a305-9d7497485d32-c000.json
-rw-r--r-- 1 hadoop hadoop 156 2022-06-15 16:11 time-country-wise-kpi/part-00045-52a168c7-5564-4842-81f1-d9c16a76d765-c000.json
-rw-r--r-- 1 hadoop hadoop 174 2022-06-15 16:10 time-country-wise-kpi/part-00049-964051e4-a8a7-47a0-9233-8645a42c8088-c000.json
-rw-r--r-- 1 hadoop hadoop 177 2022-06-15 16:14 time-country-wise-kpi/part-00065-6191ce46-d0e2-4b1e-ba50-aab4579a73fb-c000.json
-rw-r--r-- 1 hadoop hadoop 167 2022-06-15 16:17 time-country-wise-kpi/part-00073-2631e7e5-c252-42db-8912-6ff468206705-c000.json
-rw-r--r-- 1 hadoop hadoop 182 2022-06-15 16:11 time-country-wise-kpi/part-00081-41c88ae7-58a1-4d76-b583-b9f378ad239a-c000.json
-rw-r--r-- 1 hadoop hadoop 182 2022-06-15 16:10 time-country-wise-kpi/part-00081-5d66a435-714e-4bf9-bc42-c8e3e5ccdd19f-c000.json
-rw-r--r-- 1 hadoop hadoop 182 2022-06-15 16:08 time-country-wise-kpi/part-00082-59a9a162-f76f-419b-9755-664bb9003638-c000.json
-rw-r--r-- 1 hadoop hadoop 170 2022-06-15 16:16 time-country-wise-kpi/part-00086-deb0d215-50b8-4b76-9477-2c8f26f00021-c000.json
-rw-r--r-- 1 hadoop hadoop 165 2022-06-15 16:15 time-country-wise-kpi/part-00095-ded0b8be-ab41-4a41-9e60-9542f94578fb-c000.json
-rw-r--r-- 1 hadoop hadoop 157 2022-06-15 16:08 time-country-wise-kpi/part-00108-83c35027-a744-41ba-9f88-e46da4787b58-c000.json
-rw-r--r-- 1 hadoop hadoop 168 2022-06-15 16:11 time-country-wise-kpi/part-00132-c347d6b5-1232-4a36-b5ad-148d2251e3ab-c000.json
-rw-r--r-- 1 hadoop hadoop 156 2022-06-15 16:11 time-country-wise-kpi/part-00146-ed982e74-64a2-439a-b69c-ac0371366934-c000.json
-rw-r--r-- 1 hadoop hadoop 180 2022-06-15 16:18 time-country-wise-kpi/part-00153-14d8dcd9-b8ca-433d-ab49-be93600d9fc6-c000.json
-rw-r--r-- 1 hadoop hadoop 177 2022-06-15 16:13 time-country-wise-kpi/part-00171-619e513b-bd5a-47c0-b39e-4734228160ee-c000.json
-rw-r--r-- 1 hadoop hadoop 154 2022-06-15 16:11 time-country-wise-kpi/part-00175-44606e86-3c64-4fcb-80f1-6eb766f281db-c000.json
-rw-r--r-- 1 hadoop hadoop 164 2022-06-15 16:16 time-country-wise-kpi/part-00180-328b9507-76e6-4f05-b506-3c16e01d4285-c000.json
-rw-r--r-- 1 hadoop hadoop 165 2022-06-15 16:12 time-country-wise-kpi/part-00185-ffa2a99d-0618-41c2-abab-cb73865996b7-c000.json
[hadoop@ip-172-31-6-228 ~]$
```

```
[hadoop@ip-172-31-6-228 ~]$ time-wise-kpi/part-00198-21224b93-lac7-4748-806b-d39e02d0b82a-c000.json
-bash: time-wise-kpi/part-00198-21224b93-lac7-4748-806b-d39e02d0b82a-c000.json: No such file or directory
[hadoop@ip-172-31-6-228 ~]$ hadoop fs -cat time-wise-kpi/part-00133-6a735a5d-2429-4e1a-858c-e433640e3d76-c000.json/
{"window":{"start":"2022-06-15T16:04:00.000Z","end":"2022-06-15T16:05:00.000Z"},"OPM":17,"total_sales_volume":789.1,"average_transaction_size":46.41764705882353,"rate_of_return":0.058823529411764705}
[hadoop@ip-172-31-6-228 ~]$ hadoop fs -cat time-country-wise-kpi/part-00185-ffa2a99d-0618-41c2-abab-cb73865996b7-c000.json/
{"window":{"start":"2022-06-15T16:05:00.000Z","end":"2022-06-15T16:06:00.000Z"},"country":"United Kingdom","OPM":10,"total_sales_volume":214.0,"rate_of_return":0.1}
[hadoop@ip-172-31-6-228 ~]$
```

```
[hadoop@ip-172-31-6-228 ~]$ time-wise-kpi/part-00198-21224b93-lac7-4748-806b-d39e02d0b82a-c000.json
-hashi: time-wise-kpi/part-00198-21224b93-lac7-4748-806b-d39e02d0b82a-c000.json: No such file or directory
[hadoop@ip-172-31-6-228 ~]$ hadoop fs -cat time-wise-kpi/part-00139-6a738a5d-24d9-4e1a-85f0-ea3360a0d4f6-c000.json/
(*window":{"start":"2022-06-15T16:04:00.000Z","end":"2022-06-15T16:05:00.000Z"},"OPM":17,"total_sales_volume":789.1,"average_transaction_size":46.41764705882353,"rate_of_return":0.058823529411764705)
[hadoop@ip-172-31-6-228 ~]$ hadoop fs -cat time-country-wise-kpi/part-00185-f4a2a99d-0618-41c2-abab-cb7386599eb7-c000.json/
(*window":{"start":"2022-06-15T16:05:00.000Z","end":"2022-06-15T16:06:00.000Z"},"country":"United Kingdom","OPM":10,"total_sales_volume":214.0,"rate_of_return":0.1)
[hadoop@ip-172-31-6-228 ~]$ hadoop fs -get /user/hadoop/time-wise-kpi ./time-wise-kpi
[hadoop@ip-172-31-6-228 ~]$
```

The screenshot shows a WinSCP window with the local directory 'C:\Users\meta\Desktop\Data Engineering Course 4 Project-Retail Data Analyst\Prashant\Final The Final File(Actual Final)' and the remote directory '/home/hadoop/'. The local directory contains various files including 1.JPG through 11.JPG, Console-output.txt, New Text Document.txt, spark-streaming.py, and spark-streaming.py.bak. The remote directory contains 'time-country-wise-kpi', 'time-wise-kpi', and 'spark-streaming.py'. A download progress dialog is open, showing the file 'part-00180-328b9507-76ed-4f05-6506-3c16d164' being downloaded from 'C:\Users\Prashant\Final The Final File(Actual Final)' to the remote directory. The progress is at 99%.

| Name | Size | Type | Changed |
|------------------------|--------|---------------------|---------------------|
| .. | | Parent directory | 15-06-2022 21:19:41 |
| 1.JPG | 252 KB | JPG File | 15-06-2022 21:49:37 |
| 2.JPG | 267 KB | JPG File | 15-06-2022 21:49:36 |
| 3.JPG | 248 KB | JPG File | 15-06-2022 21:50:13 |
| 4.JPG | 253 KB | JPG File | 15-06-2022 21:50:31 |
| 5.JPG | 306 KB | JPG File | 15-06-2022 21:50:47 |
| 6.JPG | 40 KB | JPG File | 15-06-2022 21:51:19 |
| 7.JPG | 33 KB | JPG File | 15-06-2022 21:51:44 |
| 8.JPG | 206 KB | JPG File | 15-06-2022 21:52:19 |
| 9.JPG | 260 KB | JPG File | 15-06-2022 21:52:46 |
| 10.JPG | 63 KB | JPG File | 15-06-2022 21:54:46 |
| 11.JPG | 74 KB | JPG File | 15-06-2022 21:55:43 |
| CodeLogic.docx | 69 KB | Microsoft Office... | 15-06-2022 17:44:00 |
| Console-output.txt | 20 KB | Text Document | 25-10-2021 17:28:24 |
| New Text Document.txt | 1 KB | Text Document | 15-06-2022 21:23:27 |
| spark-streaming.py | 5 KB | Python File | 15-06-2022 21:19:02 |
| spark-streaming.py.bak | 5 KB | BAK File | 15-06-2022 21:18:38 |

| Name | Size | Changed | Rights | Owner |
|-----------------------|------|---------------------|-----------|--------|
| .. | | 15-06-2022 15:54:34 | rw-rw-r-- | root |
| time-country-wise-kpi | | 15-06-2022 16:25:25 | rw-rw-r-- | hadoop |
| time-wise-kpi | | 15-06-2022 16:25:10 | rw-rw-r-- | hadoop |
| spark-streaming.py | 5 KB | 15-06-2022 15:49:02 | rw-rw-r-- | hadoop |

The screenshot shows a WinSCP window with the local directory 'C:\Users\meta\Desktop\Data Engineering Course 4 Project-Retail Data Analyst\Prashant\Final The Final File(Actual Final)' and the remote directory '/home/hadoop/'. The local directory contains various files including 1.JPG through 12.JPG, Console-output.txt, New Text Document.txt, spark-streaming.py, and spark-streaming.py.bak. The remote directory contains 'time-country-wise-kpi', 'time-wise-kpi', and 'spark-streaming.py'. A download progress dialog is open, showing the file '14' being downloaded from 'C:\Users\Prashant\Final The Final File(Actual Final)' to the remote directory. The progress is at 20%.

| Name | Size | Type | Changed |
|------------------------|--------|---------------------|---------------------|
| .. | | Parent directory | 15-06-2022 21:56:50 |
| time-country-wise-kpi | | File folder | 15-06-2022 21:56:51 |
| 1.JPG | 252 KB | JPG File | 15-06-2022 21:49:37 |
| 2.JPG | 267 KB | JPG File | 15-06-2022 21:49:36 |
| 3.JPG | 248 KB | JPG File | 15-06-2022 21:50:13 |
| 4.JPG | 253 KB | JPG File | 15-06-2022 21:50:31 |
| 5.JPG | 306 KB | JPG File | 15-06-2022 21:50:47 |
| 6.JPG | 40 KB | JPG File | 15-06-2022 21:51:19 |
| 7.JPG | 33 KB | JPG File | 15-06-2022 21:51:44 |
| 8.JPG | 206 KB | JPG File | 15-06-2022 21:52:19 |
| 9.JPG | 260 KB | JPG File | 15-06-2022 21:52:46 |
| 10.JPG | 63 KB | JPG File | 15-06-2022 21:54:46 |
| 11.JPG | 74 KB | JPG File | 15-06-2022 21:55:43 |
| 12.JPG | 177 KB | JPG File | 15-06-2022 21:57:33 |
| CodeLogic.docx | 69 KB | Microsoft Office... | 15-06-2022 17:44:00 |
| Console-output.txt | 20 KB | Text Document | 25-10-2021 17:28:24 |
| New Text Document.txt | 1 KB | Text Document | 15-06-2022 21:23:27 |
| spark-streaming.py | 5 KB | Python File | 15-06-2022 21:19:02 |
| spark-streaming.py.bak | 5 KB | BAK File | 15-06-2022 21:18:38 |

| Name | Size | Changed | Rights | Owner |
|-----------------------|------|---------------------|-----------|--------|
| .. | | 15-06-2022 15:54:34 | rw-rw-r-- | root |
| time-country-wise-kpi | | 15-06-2022 16:25:25 | rw-rw-r-- | hadoop |
| time-wise-kpi | | 15-06-2022 16:25:10 | rw-rw-r-- | hadoop |
| spark-streaming.py | 5 KB | 15-06-2022 15:49:02 | rw-rw-r-- | hadoop |

time-country-wise-kpi - ec2-3-83-21-2.compute-1.amazonaws.com - WinSCP

Local Mark Files Commands Session Options Remote Help

ec2-3-83-21-2.compute-1.amazonaws.com X New Session

Desktop

Upload Edit Properties New Find Files

C:\Users\meta\ Desktop \Data Engineering\Course 4 Project-Retail Data Analysis\Prashant\Final The Final File\Actual Final\

| Name | Size | Type | Changed | |
|------------------------|--------|----------------------|---------------------|--|
| . | | Parent directory | 15-06-2022 21:57:55 | |
| time-country-wise-kpi | | File folder | 15-06-2022 21:56:51 | |
| time-wise-kpi | | File folder | 15-06-2022 21:57:43 | |
| 1.JPG | 252 KB | JPG File | 15-06-2022 21:49:37 | |
| 2.JPG | 267 KB | JPG File | 15-06-2022 21:49:56 | |
| 3.JPG | 248 KB | JPG File | 15-06-2022 21:50:13 | |
| 4.JPG | 253 KB | JPG File | 15-06-2022 21:50:31 | |
| 5.JPG | 308 KB | JPG File | 15-06-2022 21:50:47 | |
| 6.JPG | 40 KB | JPG File | 15-06-2022 21:51:19 | |
| 7.JPG | 33 KB | JPG File | 15-06-2022 21:51:44 | |
| 8.JPG | 206 KB | JPG File | 15-06-2022 21:52:19 | |
| 9.JPG | 200 KB | JPG File | 15-06-2022 21:53:46 | |
| 10.JPG | 63 KB | JPG File | 15-06-2022 21:54:46 | |
| 11.JPG | 74 KB | JPG File | 15-06-2022 21:55:43 | |
| 12.JPG | 177 KB | JPG File | 15-06-2022 21:57:33 | |
| 13.JPG | 179 KB | JPG File | 15-06-2022 21:57:55 | |
| CodeLogic.docx | 69 KB | Microsoft Office ... | 15-06-2022 17:44:00 | |
| Console-output.txt | 20 KB | Text Document | 25-10-2021 17:28:24 | |
| New Text Document.txt | 1 KB | Text Document | 15-06-2022 21:23:27 | |
| spark-streaming.py | 5 KB | Python File | 15-06-2022 21:18:02 | |
| spark-streaming.py.bak | 5 KB | BAK File | 15-06-2022 21:18:38 | |

0.8 of 2.39 MB in 0 of 20

/home/hadoop/time-country-wise-kpi/time-country-wise-kpi/

| Name | Size | Changed | Rights | Owner |
|---|------|---------------------|-----------|--------|
| . | | 15-06-2022 16:25:25 | rw-rw-r-- | hadoop |
| spark_metadata | | 15-06-2022 16:25:25 | rw-rw-r-- | hadoop |
| part-00000-1a4e1c2-8500-4d77-4f91-1c00b6ad9135-c000.json | 0 KB | 15-06-2022 16:25:25 | rw-rw-r-- | hadoop |
| part-00000-8b8f6ace-afbe-4441-8d2c-7f36e761d-c000.json | 0 KB | 15-06-2022 16:25:25 | rw-rw-r-- | hadoop |
| part-00000-45c70e1f-c04e-46ed-872d-c7136e761d-c000.json | 0 KB | 15-06-2022 16:25:25 | rw-rw-r-- | hadoop |
| part-00000-83ca1240-594d-4348-8885-031942b8874b-c000.json | 0 KB | 15-06-2022 16:25:25 | rw-rw-r-- | hadoop |
| part-00000-0196a6d6-8079-4d21-8a0e-081ec-c25988-c000.json | 0 KB | 15-06-2022 16:25:25 | rw-rw-r-- | hadoop |
| part-00000-831b5819-7d4e-4212-ba35-2e78a5d0a21-c000.json | 0 KB | 15-06-2022 16:25:25 | rw-rw-r-- | hadoop |
| part-00000-351980a1-3d04-4c02-9a10-118a3a0e23ed-c000.json | 0 KB | 15-06-2022 16:25:25 | rw-rw-r-- | hadoop |
| part-00000-253d4a4b-6393-40a2-8f16-e189be18f1cd-c000.json | 0 KB | 15-06-2022 16:25:25 | rw-rw-r-- | hadoop |
| part-00000-32396765-2a6d-43bb-91ee-a13d9f1e1327-c000.json | 0 KB | 15-06-2022 16:25:25 | rw-rw-r-- | hadoop |
| part-00000-a37702cb-7521-4a0c-bca3-615d463c3a3a-c000.json | 0 KB | 15-06-2022 16:25:25 | rw-rw-r-- | hadoop |
| part-00000-a724ec59-bc11-4b65-9095-33f14866f02-c000.json | 0 KB | 15-06-2022 16:25:25 | rw-rw-r-- | hadoop |
| part-00000-de17eb37-3933-4098-9379-b3945c39127-c000.json | 0 KB | 15-06-2022 16:25:25 | rw-rw-r-- | hadoop |
| part-00000-8c330707-2168-4a4d-9547-022669e29ed-c000.json | 0 KB | 15-06-2022 16:25:25 | rw-rw-r-- | hadoop |
| part-00000-45335d5a-5c90-403d-afbe-b01319f1a87-c000.json | 0 KB | 15-06-2022 16:25:25 | rw-rw-r-- | hadoop |
| part-00000-457a1a20-2143-47d3-adbb-e78562979ea-c000.json | 0 KB | 15-06-2022 16:25:25 | rw-rw-r-- | hadoop |
| part-00000-38f58776-731d-4939-b3ad-5f61c2841a61-c000.json | 1 KB | 15-06-2022 16:25:25 | rw-rw-r-- | hadoop |
| part-00012-c34afba-31b1-414a-9b7b-99173af705d1-c000.json | 1 KB | 15-06-2022 16:25:25 | rw-rw-r-- | hadoop |
| part-00019-befc236b-9480-4b5c-07bf-c1b233ada10-c000.json | 1 KB | 15-06-2022 16:25:25 | rw-rw-r-- | hadoop |
| part-00022-4c17b3e1-14ab-429f-ae64-68459a6339b-c000.json | 1 KB | 15-06-2022 16:25:25 | rw-rw-r-- | hadoop |
| part-00034-e7939ba0-8103-4b5e-a305-9d7407485d32-c000.json | 1 KB | 15-06-2022 16:25:25 | rw-rw-r-- | hadoop |
| part-00045-52a168c7-5564-4842-8111-d1c1fa76d765-c000.json | 1 KB | 15-06-2022 16:25:25 | rw-rw-r-- | hadoop |
| part-00049-94071e4-e0a7-47d0-9233-8043ac0808-c000.json | 1 KB | 15-06-2022 16:25:25 | rw-rw-r-- | hadoop |
| part-00055-0191e4b6-d0a2-4b1c-b020-aae479a7b1b-c000.json | 1 KB | 15-06-2022 16:25:25 | rw-rw-r-- | hadoop |
| part-00073-2631a7c5-c252-42ab-8012-6846d209705-c000.json | 1 KB | 15-06-2022 16:25:25 | rw-rw-r-- | hadoop |
| part-00081-5d66a435-714d-48f9-bc42-c8a35cccd19f-c000.json | 1 KB | 15-06-2022 16:25:25 | rw-rw-r-- | hadoop |
| part-00081-41c8ba7-35a1-4d76-8383-b98778d23d1-c000.json | 1 KB | 15-06-2022 16:25:25 | rw-rw-r-- | hadoop |
| part-00082-5b9a1a2-079f-479d-9755-646dab0303b-c000.json | 1 KB | 15-06-2022 16:25:25 | rw-rw-r-- | hadoop |
| part-00086-6a6d0215-59ab-4b76-9d77-5d802602021-c000.json | 1 KB | 15-06-2022 16:25:25 | rw-rw-r-- | hadoop |
| part-00095-ded88b0e-ab11-4a41-9e60-9542948378fb-c000.json | 1 KB | 15-06-2022 16:25:25 | rw-rw-r-- | hadoop |
| part-00108-49c35027-a744-41ba-9f8b-e46da4787b58-c000.json | 1 KB | 15-06-2022 16:25:25 | rw-rw-r-- | hadoop |
| part-00132-c34f6965-1232-4a39-b3ad-1a8d251e3ab-c000.json | 1 KB | 15-06-2022 16:25:25 | rw-rw-r-- | hadoop |
| part-00146-e09d1c74-64a2-439e-b05c-ec0371368934-c000.json | 1 KB | 15-06-2022 16:25:25 | rw-rw-r-- | hadoop |
| part-00153-14d8dc09-b3ca-433d-ab49-ba936009fcd-c000.json | 1 KB | 15-06-2022 16:25:25 | rw-rw-r-- | hadoop |
| part-00171-6195131b-bd5a-47c0-b396-473a22160ee-c000.json | 1 KB | 15-06-2022 16:25:25 | rw-rw-r-- | hadoop |
| part-00175-4406a608-1c44-481c-8091-64b7602314b-c000.json | 1 KB | 15-06-2022 16:25:25 | rw-rw-r-- | hadoop |
| part-00183-32b96507-76ad-495d-836c-3c16d14d251-c000.json | 1 KB | 15-06-2022 16:25:25 | rw-rw-r-- | hadoop |
| part-00185-fa2a99d-9d18-411c-adeb-c87386996a7-c000.json | 1 KB | 15-06-2022 16:25:25 | rw-rw-r-- | hadoop |

0.8 of 3.60 KB in 0 of 38

time-wise-kpi - ec2-3-83-21-2.compute-1.amazonaws.com - WinSCP

Local Mark Files Commands Session Options Remote Help

ec2-3-83-21-2.compute-1.amazonaws.com X New Session

Desktop

Upload Edit Properties New Find Files

C:\Users\meta\ Desktop \Data Engineering\Course 4 Project-Retail Data Analysis\Prashant\Final The Final File\Actual Final\

| Name | Size | Type | Changed | |
|------------------------|--------|----------------------|---------------------|--|
| . | | Parent directory | 15-06-2022 21:57:55 | |
| time-country-wise-kpi | | File folder | 15-06-2022 21:56:51 | |
| time-wise-kpi | | File folder | 15-06-2022 21:57:43 | |
| 1.JPG | 252 KB | JPG File | 15-06-2022 21:49:37 | |
| 2.JPG | 267 KB | JPG File | 15-06-2022 21:49:56 | |
| 3.JPG | 248 KB | JPG File | 15-06-2022 21:50:13 | |
| 4.JPG | 253 KB | JPG File | 15-06-2022 21:50:31 | |
| 5.JPG | 308 KB | JPG File | 15-06-2022 21:50:47 | |
| 6.JPG | 40 KB | JPG File | 15-06-2022 21:51:19 | |
| 7.JPG | 33 KB | JPG File | 15-06-2022 21:51:44 | |
| 8.JPG | 206 KB | JPG File | 15-06-2022 21:52:19 | |
| 9.JPG | 200 KB | JPG File | 15-06-2022 21:53:46 | |
| 10.JPG | 63 KB | JPG File | 15-06-2022 21:54:46 | |
| 11.JPG | 74 KB | JPG File | 15-06-2022 21:55:43 | |
| 12.JPG | 177 KB | JPG File | 15-06-2022 21:57:33 | |
| 13.JPG | 179 KB | JPG File | 15-06-2022 21:57:55 | |
| 14.JPG | 354 KB | JPG File | 15-06-2022 21:58:43 | |
| CodeLogic.docx | 69 KB | Microsoft Office ... | 15-06-2022 17:44:00 | |
| Console-output.txt | 20 KB | Text Document | 25-10-2021 17:28:24 | |
| New Text Document.txt | 1 KB | Text Document | 15-06-2022 21:23:27 | |
| spark-streaming.py | 5 KB | Python File | 15-06-2022 21:18:02 | |
| spark-streaming.py.bak | 5 KB | BAK File | 15-06-2022 21:18:38 | |

0.8 of 2.73 MB in 0 of 21

/home/hadoop/time-wise-kpi/time-wise-kpi/

| Name | Size | Changed | Rights | Owner |
|---|------|---------------------|-----------|--------|
| . | | 15-06-2022 16:25:10 | rw-rw-r-- | hadoop |
| spark_metadata | | 15-06-2022 16:25:10 | rw-rw-r-- | hadoop |
| part-00000-01903556-1736-433e-bc2b-dc033d455d4d-c000.json | 0 KB | 15-06-2022 16:25:10 | rw-rw-r-- | hadoop |
| part-00000-4c17b3e1-14ab-429f-ae64-68459a6339b-c000.json | 0 KB | 15-06-2022 16:25:10 | rw-rw-r-- | hadoop |
| part-00000-7d61a6e-c000-4e7b-a646-352905a17947-c000.json | 0 KB | 15-06-2022 16:25:10 | rw-rw-r-- | hadoop |
| part-00000-9af1306e-cf8b-4f4e-ab07-3a6cbb18185-c000.json | 0 KB | 15-06-2022 16:25:10 | rw-rw-r-- | hadoop |
| part-00003-3a9a0e7-1163-45d6-9758-285f86b2bcb-c000.json | 0 KB | 15-06-2022 16:25:10 | rw-rw-r-- | hadoop |
| part-00000-0214f0b6-3c4e-4d58-b130-3a2555fca7b-c000.json | 0 KB | 15-06-2022 16:25:10 | rw-rw-r-- | hadoop |
| part-00000-041ea273-20ce-458a-b1a1-8cdaad8d5684-c000.json | 0 KB | 15-06-2022 16:25:10 | rw-rw-r-- | hadoop |
| part-00000-044b7eed-e577-40a7-5cc0-04f8d1f12311-c000.json | 0 KB | 15-06-2022 16:25:10 | rw-rw-r-- | hadoop |
| part-00000-76ad48cc-5d6c-4592-0795-2215a89b017c-c000.json | 0 KB | 15-06-2022 16:25:10 | rw-rw-r-- | hadoop |
| part-00000-4c1963-0954-4000-5d64-47ae9d53682a-c000.json | 0 KB | 15-06-2022 16:25:10 | rw-rw-r-- | hadoop |
| part-00000-4d1e169e-a761-410c-aad8-7c39a63219f9-c000.json | 0 KB | 15-06-2022 16:25:10 | rw-rw-r-- | hadoop |
| part-00000-7073c6c6-064d-4b14-ba70-25097151913-c000.json | 0 KB | 15-06-2022 16:25:10 | rw-rw-r-- | hadoop |
| part-00000-c5e7a1e15-1c93-4028-9618-af0797041da-c000.json | 0 KB | 15-06-2022 16:25:10 | rw-rw-r-- | hadoop |
| part-00000-c57a01e7-20af-409e-aad7-3a6a3c19373-c000.json | 0 KB | 15-06-2022 16:25:10 | rw-rw-r-- | hadoop |
| part-00000-0333a4d6-4c11-4daf-e1ed-4d009ac3d8a3-c000.json | 0 KB | 15-06-2022 16:25:10 | rw-rw-r-- | hadoop |
| part-00000-df6403a5-ac5a-4340-94d4-8698ca407711-c000.json | 0 KB | 15-06-2022 16:25:10 | rw-rw-r-- | hadoop |
| part-00004-0341bc09-6842-4a98-840f-8277ead81cf-c000.json | 1 KB | 15-06-2022 16:25:10 | rw-rw-r-- | hadoop |
| part-00012-0f18a884-2b77-431c-e1e6-e051316d707c-c000.json | 1 KB | 15-06-2022 16:25:10 | rw-rw-r-- | hadoop |
| part-00034-4b4a6919-4024-4d3c-b48c-b48b3a1f0c1e-c000.json | 1 KB | 15-06-2022 16:25:10 | rw-rw-r-- | hadoop |
| part-00062-74911238-3e40-4d11-840f-3d4826e4e481-c000.json | 1 KB | 15-06-2022 16:25:10 | rw-rw-r-- | hadoop |
| part-00087-80feda25-b743-801b-811e-0262051dec0f-c000.json | 1 KB | 15-06-2022 16:25:10 | rw-rw-r-- | hadoop |
| part-00088-b73849a3-c857-4253-ba72-79090602443-c000.json | 1 KB | 15-06-2022 16:25:10 | rw-rw-r-- | hadoop |
| part-00089-ba17a5d6-0773-4d20-b7d5-e7626a6e9b71-c000.json | 1 KB | 15-06-2022 16:25:10 | rw-rw-r-- | hadoop |
| part-00106-5b2ab871-4e4c-4e1a-858c-e4336a0c3d76-c000.json | 1 KB | 15-06-2022 16:25:10 | rw-rw-r-- | hadoop |
| part-00133-4a735a5d-2429-4e1a-858c-e4336a0c3d76-c000.json | 1 KB | 15-06-2022 16:25:10 | rw-rw-r-- | hadoop |
| part-00154-203a6b62-79ec-4c10-8058-6d561be2a84a-c000.json | 1 KB | 15-06-2022 16:25:10 | rw-rw-r-- | hadoop |
| part-00189-21246893-1ac7-4748-806b-d39e3d0e82a-c000.json | 1 KB | 15-06-2022 16:25:10 | rw-rw-r-- | hadoop |

0.8 of 2.09 KB in 0 of 28