

## SETUP [ on CDH + Spark2 ]

Setting Environment before running the Spark Job:

```
export SPARK_KAFKA_VERSION=0.10
```

Downloaded the Jar file using:

```
wget https://ds-spark-sql-kafka-jar.s3.amazonaws.com/spark-sql-kafka-0-10\_2.11-2.3.0.jar
```

Commands used to run the Spark Submit job:

```
spark2-submit --jars spark-sql-kafka-0-10_2.11-2.3.0.jar spark-streaming.py
```

## UTILITY FUNCTIONS

**total\_item\_count**: to sum up quantity of items ordered for each invoice

**total\_cost**: get total cost using  $\rightarrow \text{quantity} * \text{unit\_price}$  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

## 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.

Output HDFS Dir for agg\_time: **timeKPI/**

Output HDFS Dir for agg\_time\_country: **time\_countryKPI/**

## CONSOLE OUTPUT

Batch: 5

invoice_no	country	timestamp	type	total_items	total_cost	is_order	is_return
154132545944339	United Kingdom	2021-09-13 08:19:28	ORDER	54	61.79	1	0
154132545944340	France	2021-09-13 08:19:44	ORDER	15	18.6	1	0
154132545944341	United Kingdom	2021-09-13 08:19:48	ORDER	6	76.5	1	0
154132545944342	United Kingdom	2021-09-13 08:20:04	ORDER	2	6.2	1	0
154132545944343	Germany	2021-09-13 08:20:12	ORDER	9	17.49	1	0

Batch: 6

invoice_no	country	timestamp	type	total_items	total_cost	is_order	is_return
154132545944344	United Kingdom	2021-09-13 08:20:24	ORDER	135	180.23	1	0
154132545944345	United Kingdom	2021-09-13 08:20:29	ORDER	3	2.49	1	0
154132545944346	United Kingdom	2021-09-13 08:20:41	ORDER	20	35.06	1	0
154132545944347	United Kingdom	2021-09-13 08:20:48	ORDER	39	48.91	1	0
154132545944348	United Kingdom	2021-09-13 08:20:49	ORDER	12	23.400002	1	0
154132545944349	United Kingdom	2021-09-13 08:20:51	ORDER	9	37.6	1	0
154132545944350	France	2021-09-13 08:20:51	ORDER	31	97.67	1	0
154132545944351	United Kingdom	2021-09-13 08:20:56	ORDER	97	290.21	1	0
154132545944352	United Kingdom	2021-09-13 08:21:13	ORDER	28	80.56	1	0

Batch: 7

invoice_no	country	timestamp	type	total_items	total_cost	is_order	is_return
154132545944353	United Kingdom	2021-09-13 08:21:37	ORDER	12	39.48	1	0
154132545944354	United Kingdom	2021-09-13 08:21:42	ORDER	4	33.4	1	0
154132545944355	United Kingdom	2021-09-13 08:21:45	ORDER	49	81.670006	1	0
154132545944356	United Kingdom	2021-09-13 08:21:48	ORDER	29	152.89	1	0
154132545944357	United Kingdom	2021-09-13 08:21:53	ORDER	19	48.309998	1	0