

Load data from AWS RDS to Hadoop

<Command to run the python file>

export SPARK_KAFKA_VERSION=0.10

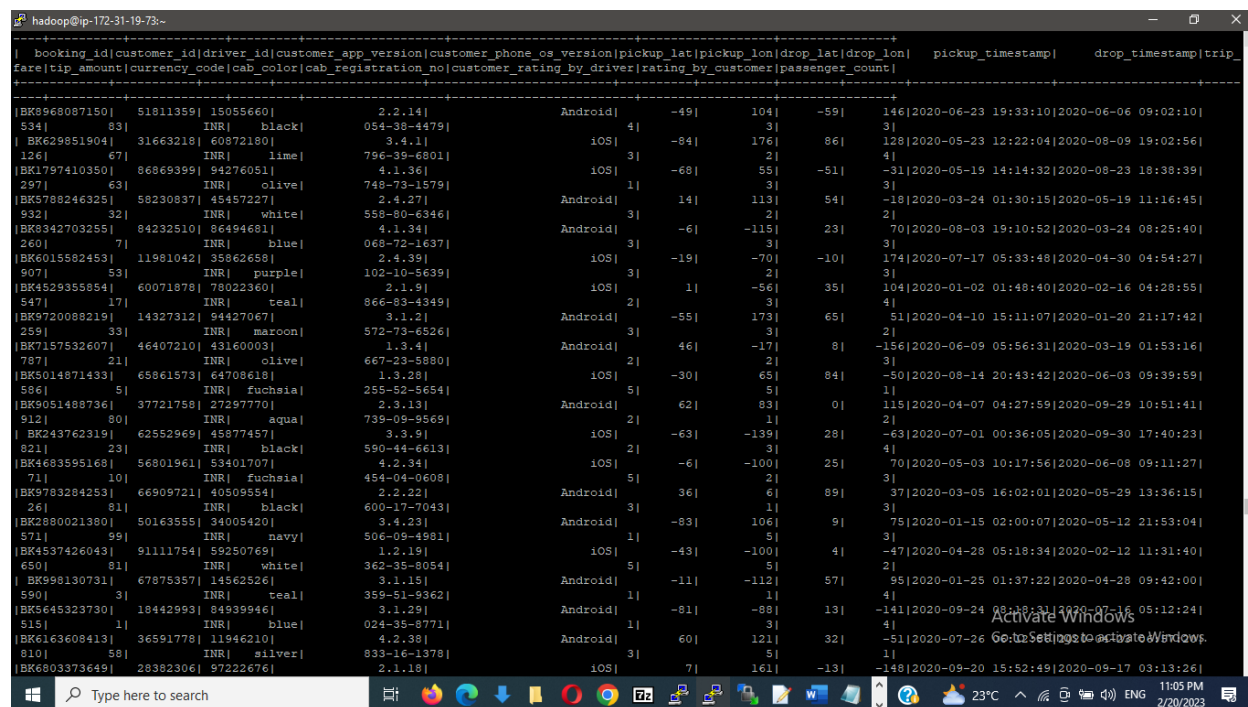
spark-submit --packages org.apache.spark:spark-sql-kafka-0-10_2.11:2.4.5

datewise_bookings_aggregates_spark.py

<Command to move the csv file to HDFS>

hadoop fs -put /home/hadoop/datewise_booking /user/capstone/carride2/aggregated1/

<Screenshot of the file in HDFS>



booking_id	customer_id	driver_id	customer_app_version	customer_phone_os_version	pickup_lat	pickup_lon	drop_lat	drop_lon	pickup_timestamp	drop_timestamp	trip_fare
534	83	15811359	15055660	2.2.14	Android	-49	104	-59	146 2020-06-23 19:33:10	2020-06-06 09:02:10	3
126	67	31663218	60872180	3.4.1	iOS	-84	176	86	128 2020-05-23 12:22:04	2020-08-09 19:02:56	4
297	63	86869399	94276051	4.1.36	iOS	-68	55	-51	-31 2020-05-19 14:14:32	2020-08-23 10:38:39	3
932	32	58230837	45457227	2.4.27	Android	14	113	54	-18 2020-03-24 01:30:15	2020-05-19 11:16:45	2
260	7	84232510	86494681	4.1.34	Android	-6	-115	23	70 2020-08-03 19:10:52	2020-03-24 08:25:40	3
907	53	11981042	35862658	2.4.39	iOS	-19	-70	-10	174 2020-07-17 05:33:48	2020-04-30 04:54:27	3
547	17	60071878	78022360	2.1.91	iOS	1	-56	35	104 2020-01-02 01:48:40	2020-02-16 04:28:55	4
787	21	14327312	94427067	3.1.2	Android	-55	173	65	51 2020-04-10 15:11:07	2020-01-20 21:17:42	2
586	5	46407210	43160003	1.3.4	Android	46	-17	8	-156 2020-06-09 05:56:31	2020-03-19 01:53:16	3
912	30	65861573	64708618	1.3.28	iOS	-30	65	84	-50 2020-08-14 20:43:42	2020-06-03 09:39:59	1
821	23	37721758	27297770	2.3.13	Android	62	83	0	115 2020-04-07 04:27:59	2020-09-29 10:51:41	2
71	10	62552969	45877457	3.3.9	iOS	-63	-139	28	-63 2020-07-01 00:36:05	2020-09-30 17:40:23	4
26	81	56801961	53401707	4.2.34	iOS	-6	-100	25	70 2020-05-03 10:17:56	2020-06-08 09:11:27	3
571	99	66909721	40509554	2.2.22	Android	36	61	89	37 2020-03-05 16:02:01	2020-05-29 13:36:15	3
650	81	50163555	34005420	3.4.23	Android	-83	106	9	75 2020-01-15 02:00:07	2020-05-12 21:53:04	3
590	3	91111754	59250769	1.2.19	iOS	-43	-100	4	-47 2020-04-28 05:18:34	2020-02-12 11:31:40	2
515	1	67875357	14562526	3.1.15	Android	-11	-112	57	95 2020-01-25 01:37:22	2020-04-28 09:42:00	4
810	58	18442993	84939946	3.1.29	Android	-81	-88	13	-141 2020-09-24 08:18:31	2020-07-16 05:12:24	4
534	83	36591778	11946210	4.2.38	Android	60	121	32	-51 2020-07-26 06:19:31	2020-07-16 05:12:24	4
126	67	28382306	97222676	2.1.18	iOS	7	161	-13	-148 2020-09-20 15:52:49	2020-09-17 03:13:26	1

```
23/02/20 17:31:21 INFO DAGScheduler: Job 3 finished: showST1
+-----+-----+
|      pickup_date|total_bookings|
+-----+-----+
|2020-03-31 00:00:00|          6|
|2020-08-19 00:00:00|          2|
|2020-03-15 00:00:00|          3|
|2020-08-23 00:00:00|          3|
|2020-02-07 00:00:00|          3|
|2020-06-14 00:00:00|          4|
|2020-02-20 00:00:00|          3|
|2020-02-12 00:00:00|          3|
|2020-04-06 00:00:00|          1|
|2020-09-10 00:00:00|          3|
|2020-06-22 00:00:00|          2|
|2020-02-06 00:00:00|          3|
|2020-05-20 00:00:00|          3|
|2020-09-18 00:00:00|          2|
|2020-08-18 00:00:00|          5|
|2020-02-24 00:00:00|          6|
|2020-02-23 00:00:00|          1|
|2020-10-22 00:00:00|          3|
|2020-10-16 00:00:00|          4|
|2020-04-29 00:00:00|          5|
+-----+-----+
only showing top 20 rows
```