

Advanced Topics in Database Systems

Semester Assignment

Group 19

Vlachakis Nikos, el18441

Apostolos Garos, el18198

Github Link: https://github.com/ApostolisGaros/Spark_Hadoop_Project_AdvancedDB

1) Q1 time taken: 12.387681245803833 seconds. (2 workers)

Q1 time taken: 13.106610298156738 seconds. (1 worker)

Attributes	Values
VendorID	2
tpep_pickup_datetime	2022-03-17 12:27:47
tpep_dropoff_datetime	2022-03-17 12:27:58
passenger_count	1.0
trip_distance	0.0
RatecodeID	1.0
store_and_fwd_flag	N
PULocationID	12
DOLocationID	12
payment_type	1
fare_amount	2.5
extra	0.0
mta_tax	0.5
tip_amount	40.0
tolls_amount	0.0
improvement_surcharge	0.3
total_amount	45.8
congestion_surcharge	2.5
airport_fee	0.0

2) Q2 time taken: 35.75651144981384 seconds. (2 workers)

Q2 time taken: 36.10699152946472 seconds. (1 worker)

Attributes	January	March	July	May	February	October	April	June
VendorID	1	1	2	1	1	2	1	1
tpep_pickup_datetime	2022-01-22 11:39:07	2022-03-11 20:08:32	2022-07-01 00:43:38	2022-05-21 16:47:48	2022-02-18 02:33:30	2002-10-25 00:33:28	2022-04-29 04:31:21	2022-06-12 16:51:46
tpep_dropoff_datetime	2022-01-22 12:31:09	2022-03-11 20:09:45	2022-07-01 01:35:51	2022-05-21 17:05:47	2022-02-18 02:35:28	2002-10-25 00:35:16	2022-04-29 04:32:30	2022-06-12 17:56:48
passenger_count	1.0	1.0	1.0	1.0	1.0	2.0	2.0	9.0
trip_distance	33.4	0.0	32.58	2.4	1.3	0.0	0.0	22.0
RatecodeID	1.0	1.0	3.0	3.0	1.0	5.0	1.0	1.0
store_and_fwd_flag	Y	N	N	N	N	N	N	N
PULocationID	70	265	138	239	265	265	249	142
DOLocationID	265	265	1	246	265	265	249	132
payment_type	4	1	1	3	1	2	3	2
fare_amount	88.0	2.5	104.5	31.5	3.0	100.0	3.0	67.5
extra	0.0	1.0	0.5	0.0	0.5	0.0	3.0	2.5
mta_tax	0.5	0.5	0.0	0.0	0.5	0.0	0.5	0.5
tip_amount	0.0	48.0	33.71	0.0	19.85	0.0	0.0	0.0
tolls_amount	193.3	235.7	29.55	813.75	95.0	18.3	911.87	800.09
improvement_surcharge	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3
total_amount	282.1	288.0	169.81	845.55	119.15	118.6	918.67	870.89
congestion_surcharge	0.0	0.0	0.0	0.0	0.0	0.0	2.5	2.5
tip_percentage	0.0	0.0	1.25	0.0	0.0	0.0	0.0	0.0

3) Q3_DF time taken: 21.829453229904175 seconds.(2 workers)

Q3_DF time taken: 21.71215009689331 seconds. (1 worker)

Q3_RDD time taken: 323.1897180080414 seconds. (2 workers)

Q3_RDD time taken: 321.54228043556213 seconds. (1 worker)

group	15_day_avg_trip_distance	15_day_avg_total_amount
0	5.58	20.04
1	5.23	19.02
2	6.0	19.6
3	6.17	20.2
4	6.69	20.72
5	5.51	21.15
6	5.67	21.54
7	5.81	21.43
8	6.26	21.94
9	7.89	22.79
10	6.38	22.46
11	6.16	22.36
12	4.89	22.45

- 4) Q4 time taken: 18.941542863845825 seconds. (2 workers)
Q4 time taken: 18.249610662460327 seconds. (1 worker)

dayofweek(tpep_pickup_datetime)	hour(tpep_pickup_datetime)	max_passenger_count	index
1	16	9.0	1
1	20	8.0	2
1	22	8.0	3
2	17	9.0	1
2	1	9.0	2
2	20	9.0	3
3	22	9.0	1
3	20	9.0	2
3	19	9.0	3
4	9	9.0	1
4	20	8.0	2
4	21	8.0	3
5	21	9.0	1
5	2	9.0	2
5	17	8.0	3
6	18	9.0	1
6	19	9.0	2
6	20	8.0	3
7	14	9.0	1
7	13	9.0	2

- 5) Q5 time taken: 19.15152072906494 seconds. (2 workers)
 Q5 time taken: 18.670414924621582 seconds. (1 worker)

month(tpep_pickup_datetime)	dayofmonth(tpep_pickup_datetime)	tip_percentage	index
1	29	0.215483366934815	1
1	15	0.19532266162943906	2
1	22	0.1933725570230018	3
1	30	0.19280650021908916	4
1	21	0.19276749580993513	5
2	4	0.195575922031784	1
2	5	0.19534122595795528	2
2	6	0.19400636200035978	3
2	10	0.193551754296804	4
2	17	0.1929098458334977	5
3	9	0.19555776092383564	1
3	12	0.1939201787965798	2
3	30	0.1932932451137539	3
3	24	0.19278501075241727	4
3	10	0.192740557900796	5
4	1	0.1913783343379289	1
4	7	0.1912479413453781	2
4	6	0.1909122089695774	3
4	27	0.19031992584857682	4
4	28	0.18936868740705726	5
5	12	0.192141622662075	1
5	4	0.19138776780086114	2
5	11	0.190290914716498	3
5	10	0.18971889275613088	4
5	6	0.18968117489605255	5
6	16	0.190438716389348	1
6	8	0.18967537927853564	2
6	23	0.18921619118071123	3
6	9	0.18911257641776286	4
6	17	0.18829881578836138	5
7	1	0.19232859944686387	1
8	23	0.0	1
10	27	0.09431522570772764	1
10	23	0.0880983750548968	2
10	25	0.07683164327446107	3
10	26	0.0594137127437753	4
10	21	0.05834004024144869	5
12	31	0.1111111111111111	1