



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ
ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΠΡΟΧΩΡΗΜΕΝΑ ΘΕΜΑΤΑ ΒΑΣΕΩΝ ΔΕΔΟΜΕΝΩΝ

ΧΕΙΜΕΡΙΝΟ ΕΞΑΜΗΝΟ 2022-2023

ΟΜΑΔΑ 44:
ΚΟΚΚΙΝΑΚΗΣ ΠΑΝΑΓΙΩΤΗΣ 03118115
ΚΟΝΤΑΛΕΞΗ ΜΑΡΙΝΑ 03118022

Το github repository της εργασίας είναι το https://github.com/marinakontalexi/advDB_2022.

Χρησιμοποιούμε ένα Worker Node στον Master και ένα στον Slave όταν θέλουμε να τρέξουμε τα queries με 2 workers.

2.

Q1:

	1 Worker Node	2 Worker Nodes
Time (s)	12.928159952163696	12.47849988937378

Result - Max Tip: 40\$

Ο πίνακας που περιέχει τα πλήρη στοιχεία της απάντησης του Q1 βρίσκεται στο αρχείο *q1_result.csv* που βρίσκεται στο github repository της εργασίας.

Q2:

	1 Worker	2 Workers
Time(s)	76.92775678634644	97.57905864715576

Result:

Month	Max Tolls
1	193.3
2	95.0
3	235.7
4	911.87
5	813.75
6	800.09

Ο πίνακας που περιέχει τα πλήρη στοιχεία της απάντησης του Q2 βρίσκεται στο αρχείο *q2_result.csv* που βρίσκεται στο github repository της εργασίας.

3.

Q3 SQL

	1 Worker	2 Workers
Time(s)	22.409590482711792	21.484654903411865

Q3 RDD

	1 Worker	2 Workers
Time(s)	294.52160859	284.0756914615631

Result

Fortnight	Average_Distance	Average_Total_Fare
0	5.576429554927404	19.90405084638674
1	5.097880367275346	19.14882164234129
2	6.248888338463885	19.491979067238955
3	5.849460516243601	20.18769180439039
4	6.480485434052913	20.652278174141436
5	5.556944935850653	21.120920554171548
6	5.6793230779383554	21.515559094570637
7	5.8003418315344	21.428117467105963
8	6.249697852127271	21.92157034889687
9	6.315157336730044	22.466305309372895
10	7.906694182348759	22.771948777963715
11	6.174138574511356	22.331380641103525

4.
Q4:

	1 Worker	2 Workers
Time(s)	24.31680727005005	25.063889026641846

Result:

Weekday	Hour
1	0
1	19
1	17
2	20
2	21
2	19
3	20
3	21
3	19
4	20
4	21
4	19
5	20
5	21
5	19
6	21
6	20
6	22
7	21
7	20
7	19

Q5:

	1 Worker	2 Workers
Time(s)	20.055628776550293	19.715713262557983

Result:

Month	Day Of Month
1	9
1	31
1	1
1	29
1	16
2	21
2	13
2	9
2	10
2	27
3	18
3	21
3	26
3	5
3	12
4	12
4	2
4	21
4	3
4	30
5	12
5	20
5	16

5	15
5	6
6	13
6	25
6	10
6	16
6	20

Plot χρόνων εκτέλεσης για 1 και 2 Workers ανά SQL Query:

