

## Συστήματα Διαχείρισης Δεδομένων Μεγάλου Όγκου

### Εργαστηριακή Άσκηση 2022/23

Όνομα	Επώνυμο	ΑΜ
Αριάδνη	Μπαρμπαρή	1067379
Ελένη	Τασόκα	1067530

Βεβαιώνω ότι είμαι συγγραφέας της παρούσας εργασίας και ότι έχω αναφέρει ή παραπέμψει σε αυτήν, ρητά και συγκεκριμένα, όλες τις πηγές από τις οποίες έκανα χρήση δεδομένων, ιδεών, προτάσεων ή λέξεων, είτε αυτές μεταφέρονται επακριβώς (στο πρωτότυπο ή μεταφρασμένες) είτε παραφρασμένες. Επίσης βεβαιώνω ότι αυτή η εργασία προετοιμάστηκε από εμένα προσωπικά ειδικά για το συγκεκριμένο μάθημα/σεμινάριο/πρόγραμμα σπουδών.

Έχω ενημερωθεί ότι σύμφωνα με τον εσωτερικό κανονισμό λειτουργίας του Πανεπιστημίου Πατρών άρθρο 50§6, τυχόν προσπάθεια αντιγραφής ή εν γένει φалκίδευσης της εξεταστικής και εκπαιδευτικής διαδικασίας από οιονδήποτε εξεταζόμενο, πέραν του μηδενισμού, συνιστά βαρύ πειθαρχικό παράπτωμα.

Υπογραφή

Μπαρμπαρή Αριάδνη

22 / 09 / 2023

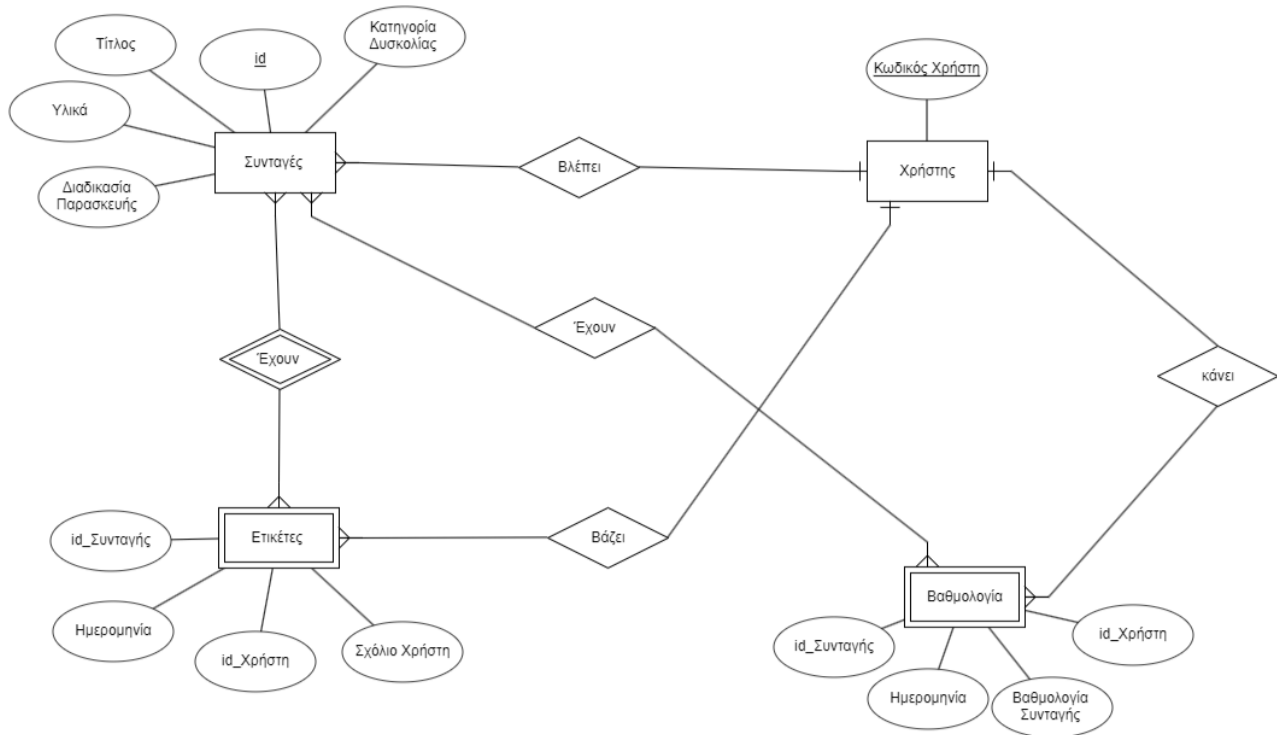
Υπογραφή

Τασόκα Ελένη

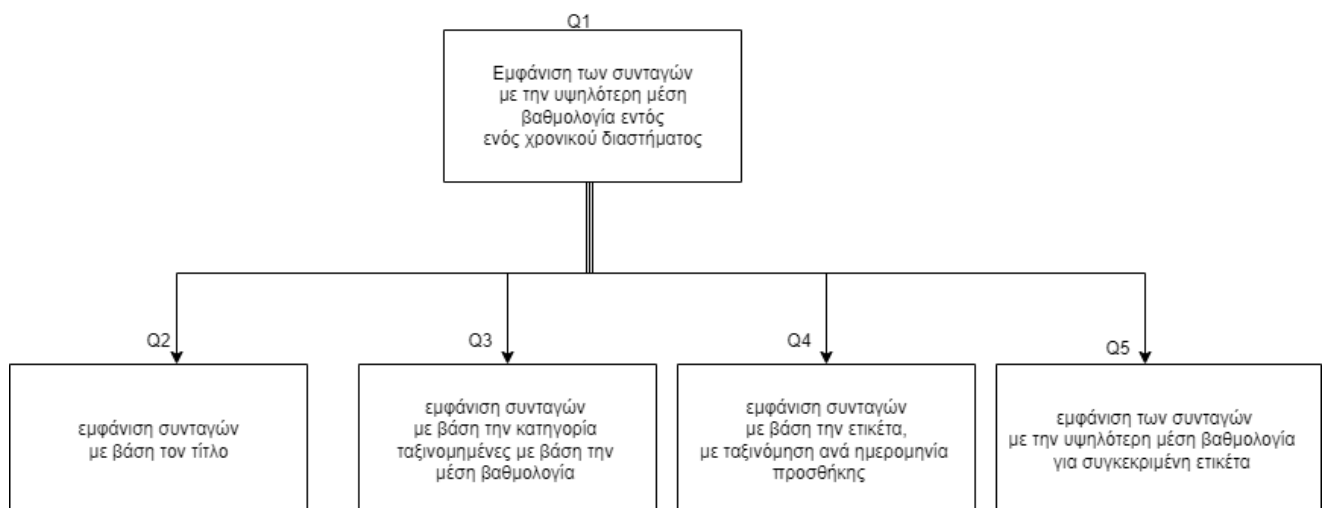
22/ 09 / 2023

## Ερώτημα 1: Σχεδιασμός ΒΔ

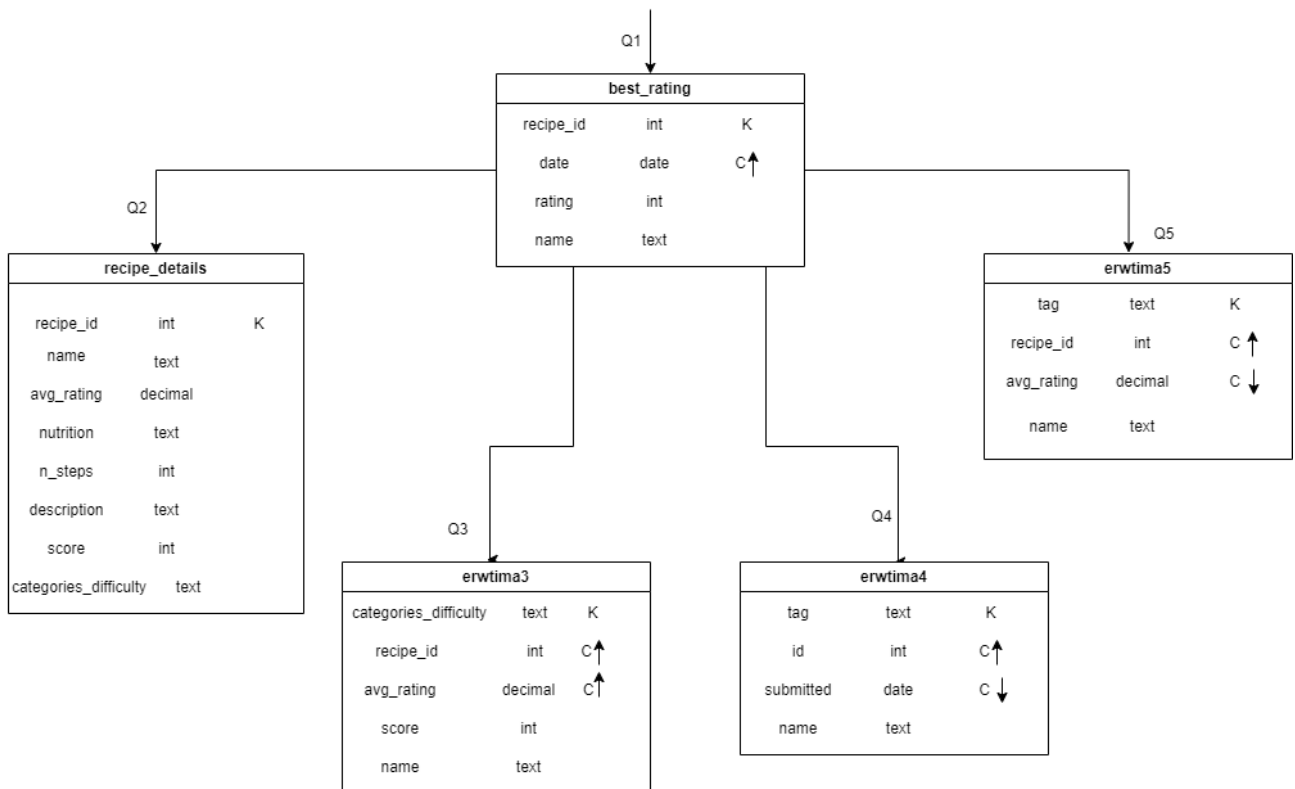
### Εννοιολογικό Μοντέλο



### Application Workflow



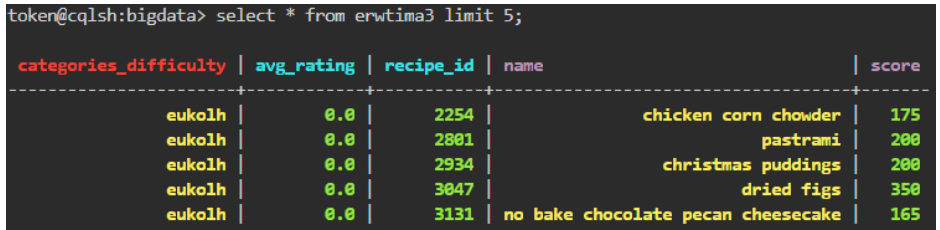
## Chebotko Diagram



## Ερώτημα 2: Ερωτήματα DDL

Keyspace	best_rating																								
DDL statement	<pre>#create table session.execute ("""CREATE TABLE IF NOT EXISTS best_rating (     recipe_id int,     date date,     rating int,     name text,     PRIMARY KEY ((recipe_id), date) ) WITH CLUSTERING ORDER BY (date ASC); """)</pre>																								
Screenshot	<pre>token@cqlsh&gt; use bigdata; token@cqlsh:bigdata&gt; select * from best_rating limit 5;</pre> <table><tr><th>recipe_id</th><th>date</th><th>name</th><th>rating</th></tr><tr><td>265902</td><td>2008-03-16</td><td>sausage bean and cabbage soup</td><td>5</td></tr><tr><td>265902</td><td>2011-01-31</td><td>sausage bean and cabbage soup</td><td>5</td></tr><tr><td>176996</td><td>2008-09-28</td><td>0 fat chunky watermelon salsa</td><td>5</td></tr><tr><td>143843</td><td>2007-11-16</td><td>almond ginger biscotti</td><td>5</td></tr><tr><td>77328</td><td>2003-12-19</td><td>best double chocolate muffins</td><td>3</td></tr></table>	recipe_id	date	name	rating	265902	2008-03-16	sausage bean and cabbage soup	5	265902	2011-01-31	sausage bean and cabbage soup	5	176996	2008-09-28	0 fat chunky watermelon salsa	5	143843	2007-11-16	almond ginger biscotti	5	77328	2003-12-19	best double chocolate muffins	3
recipe_id	date	name	rating																						
265902	2008-03-16	sausage bean and cabbage soup	5																						
265902	2011-01-31	sausage bean and cabbage soup	5																						
176996	2008-09-28	0 fat chunky watermelon salsa	5																						
143843	2007-11-16	almond ginger biscotti	5																						
77328	2003-12-19	best double chocolate muffins	3																						

Keyspace	recipe_details																																																
DDL statement	<pre>#create table session.execute ("""CREATE TABLE IF NOT EXISTS recipe_details (     recipe_id int,     name text,     avg_rating decimal,     nutrition text,     n_steps int,     description text,     score int,     categories_difficulty text,     PRIMARY KEY (recipe_id) ) """)</pre>																																																
Screenshot	 <pre>token@cqlsh:bigdata&gt; select * from recipe_details limit 5;</pre> <table><thead><tr><th>recipe_id</th><th>avg_rating</th><th>categories_difficulty</th><th>description</th><th>n_steps</th><th>name</th><th>nutrition</th><th>score</th></tr></thead><tbody><tr><td>265902</td><td>5.0</td><td>poly eukolh</td><td>found this newspaper clipping tucked away among some old papers i was searching through but dont know what newspaper it came from sounds perfect for these cold nights</td><td>9</td><td>sausage bean and cabbage soup</td><td>2689 160 210 490 220 130 110</td><td>100</td></tr><tr><td>176996</td><td>5.0</td><td>metria</td><td>this is fantastic served with grilled seafood or chicken plan ahead this needs to chill for 30 minutes before serving</td><td>6</td><td>fat chunky watermelon salsa</td><td>577 00 390 00 20 00 40</td><td>455</td></tr><tr><td>143843</td><td>5.0</td><td>poly eukolh</td><td>from everyday food october 2005 allow extra time for cooling</td><td>16</td><td>almond ginger biscotti</td><td>829 50 180 10 50 10 30</td><td>60</td></tr><tr><td>77328</td><td>3.0</td><td>metria</td><td>the absolute best chocolate muffin recipe aroundyou wont just stop at eating one of thesealthough this recipe makes 12 muffins better make a double batch in two separate bowls cause you will need it</td><td>8</td><td>best double chocolate muffins</td><td>2548 230 340 130 100 340 90</td><td>550</td></tr><tr><td>296370</td><td>5.0</td><td>eukolh</td><td>a great way to use up the bag of turnips the helpful neighbor lady gave you</td><td>19</td><td>orange kissed turnip custards</td><td>1104 20 480 00 180 10 50</td><td>185</td></tr></tbody></table>	recipe_id	avg_rating	categories_difficulty	description	n_steps	name	nutrition	score	265902	5.0	poly eukolh	found this newspaper clipping tucked away among some old papers i was searching through but dont know what newspaper it came from sounds perfect for these cold nights	9	sausage bean and cabbage soup	2689 160 210 490 220 130 110	100	176996	5.0	metria	this is fantastic served with grilled seafood or chicken plan ahead this needs to chill for 30 minutes before serving	6	fat chunky watermelon salsa	577 00 390 00 20 00 40	455	143843	5.0	poly eukolh	from everyday food october 2005 allow extra time for cooling	16	almond ginger biscotti	829 50 180 10 50 10 30	60	77328	3.0	metria	the absolute best chocolate muffin recipe aroundyou wont just stop at eating one of thesealthough this recipe makes 12 muffins better make a double batch in two separate bowls cause you will need it	8	best double chocolate muffins	2548 230 340 130 100 340 90	550	296370	5.0	eukolh	a great way to use up the bag of turnips the helpful neighbor lady gave you	19	orange kissed turnip custards	1104 20 480 00 180 10 50	185
recipe_id	avg_rating	categories_difficulty	description	n_steps	name	nutrition	score																																										
265902	5.0	poly eukolh	found this newspaper clipping tucked away among some old papers i was searching through but dont know what newspaper it came from sounds perfect for these cold nights	9	sausage bean and cabbage soup	2689 160 210 490 220 130 110	100																																										
176996	5.0	metria	this is fantastic served with grilled seafood or chicken plan ahead this needs to chill for 30 minutes before serving	6	fat chunky watermelon salsa	577 00 390 00 20 00 40	455																																										
143843	5.0	poly eukolh	from everyday food october 2005 allow extra time for cooling	16	almond ginger biscotti	829 50 180 10 50 10 30	60																																										
77328	3.0	metria	the absolute best chocolate muffin recipe aroundyou wont just stop at eating one of thesealthough this recipe makes 12 muffins better make a double batch in two separate bowls cause you will need it	8	best double chocolate muffins	2548 230 340 130 100 340 90	550																																										
296370	5.0	eukolh	a great way to use up the bag of turnips the helpful neighbor lady gave you	19	orange kissed turnip custards	1104 20 480 00 180 10 50	185																																										

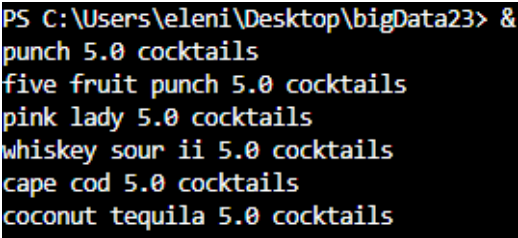
Keyspace	erwtima3																														
DDL statement	<pre>#create table session.execute ("""CREATE TABLE IF NOT EXISTS erwtima3 (     recipe_id int,     name text,     avg_rating decimal,     score int,     categories_difficulty text,     PRIMARY KEY ((categories_difficulty), avg_rating, recipe_id) ) WITH CLUSTERING ORDER BY (avg_rating ASC, recipe_id ASC); """)</pre>																														
Screenshot	 <pre>token@cqlsh:bigdata&gt; select * from erwtima3 limit 5;</pre> <table><thead><tr><th>categories_difficulty</th><th>avg_rating</th><th>recipe_id</th><th>name</th><th>score</th></tr></thead><tbody><tr><td>eukolh</td><td>0.0</td><td>2254</td><td>chicken corn chowder</td><td>175</td></tr><tr><td>eukolh</td><td>0.0</td><td>2801</td><td>pastrami</td><td>200</td></tr><tr><td>eukolh</td><td>0.0</td><td>2934</td><td>christmas puddings</td><td>200</td></tr><tr><td>eukolh</td><td>0.0</td><td>3047</td><td>dried figs</td><td>350</td></tr><tr><td>eukolh</td><td>0.0</td><td>3131</td><td>no bake chocolate pecan cheesecake</td><td>165</td></tr></tbody></table>	categories_difficulty	avg_rating	recipe_id	name	score	eukolh	0.0	2254	chicken corn chowder	175	eukolh	0.0	2801	pastrami	200	eukolh	0.0	2934	christmas puddings	200	eukolh	0.0	3047	dried figs	350	eukolh	0.0	3131	no bake chocolate pecan cheesecake	165
categories_difficulty	avg_rating	recipe_id	name	score																											
eukolh	0.0	2254	chicken corn chowder	175																											
eukolh	0.0	2801	pastrami	200																											
eukolh	0.0	2934	christmas puddings	200																											
eukolh	0.0	3047	dried figs	350																											
eukolh	0.0	3131	no bake chocolate pecan cheesecake	165																											

Keyspace	erwtima4																								
DDL statement	<pre>#create table session.execute ("""CREATE TABLE IF NOT EXISTS erwtima4     name text,     id int,     tag text,     submitted date,     PRIMARY KEY ((tag), submitted,id)   ) WITH CLUSTERING ORDER BY (submitted DESC, id ASC);   """)</pre>																								
Screenshot	<pre>token@cqlsh:bigdata&gt; select * from erwtima4 limit 5;</pre> <table><thead><tr><th>tag</th><th>submitted</th><th>id</th><th>name</th></tr></thead><tbody><tr><td>low cholesterol</td><td>2018-09-04</td><td>536859</td><td>spicy red pepper jelly</td></tr><tr><td>low cholesterol</td><td>2018-08-12</td><td>536734</td><td>grilled cajun green beans</td></tr><tr><td>low cholesterol</td><td>2018-08-11</td><td>536729</td><td>creole watermelon feta salad</td></tr><tr><td>low cholesterol</td><td>2018-07-30</td><td>536547</td><td>cauliflower ceviche</td></tr><tr><td>low cholesterol</td><td>2018-07-24</td><td>536443</td><td>grilled beet charred citrus salad</td></tr></tbody></table>	tag	submitted	id	name	low cholesterol	2018-09-04	536859	spicy red pepper jelly	low cholesterol	2018-08-12	536734	grilled cajun green beans	low cholesterol	2018-08-11	536729	creole watermelon feta salad	low cholesterol	2018-07-30	536547	cauliflower ceviche	low cholesterol	2018-07-24	536443	grilled beet charred citrus salad
tag	submitted	id	name																						
low cholesterol	2018-09-04	536859	spicy red pepper jelly																						
low cholesterol	2018-08-12	536734	grilled cajun green beans																						
low cholesterol	2018-08-11	536729	creole watermelon feta salad																						
low cholesterol	2018-07-30	536547	cauliflower ceviche																						
low cholesterol	2018-07-24	536443	grilled beet charred citrus salad																						

Keyspace	erwtima5																								
DDL statement	<pre>#create table session.execute ("""CREATE TABLE IF NOT EXISTS erwtima5 (     recipe_id int,     name text,     avg_rating decimal,     tag text,     PRIMARY KEY ((tag), avg_rating,recipe_id) ) WITH CLUSTERING ORDER BY (avg_rating DESC, recipe_id ASC); """)</pre>																								
Screenshot	<pre>token@cqlsh:bigdata&gt; select * from erwtima5 limit 5;</pre> <table><thead><tr><th>tag</th><th>avg_rating</th><th>recipe_id</th><th>name</th></tr></thead><tbody><tr><td>low cholesterol</td><td>5.0</td><td>52</td><td>cafe cappuccino</td></tr><tr><td>low cholesterol</td><td>5.0</td><td>59</td><td>lou s fabulous bruschetta</td></tr><tr><td>low cholesterol</td><td>5.0</td><td>128</td><td>caramel apple milkshakes</td></tr><tr><td>low cholesterol</td><td>5.0</td><td>135</td><td>cherry sandwich maker snack</td></tr><tr><td>low cholesterol</td><td>5.0</td><td>147</td><td>campfire orange cake</td></tr></tbody></table>	tag	avg_rating	recipe_id	name	low cholesterol	5.0	52	cafe cappuccino	low cholesterol	5.0	59	lou s fabulous bruschetta	low cholesterol	5.0	128	caramel apple milkshakes	low cholesterol	5.0	135	cherry sandwich maker snack	low cholesterol	5.0	147	campfire orange cake
tag	avg_rating	recipe_id	name																						
low cholesterol	5.0	52	cafe cappuccino																						
low cholesterol	5.0	59	lou s fabulous bruschetta																						
low cholesterol	5.0	128	caramel apple milkshakes																						
low cholesterol	5.0	135	cherry sandwich maker snack																						
low cholesterol	5.0	147	campfire orange cake																						

### Ερώτημα 3: Απαντήσεις ερωτημάτων

Ερώτημα	Απάντηση
Εμφάνιση των 30 συνταγών με την υψηλότερη μέση βαθμολογία μεταξύ 01/01/2012 και 31/05/2012	<pre>PS C:\Users\eleni\Desktop\bigData23&gt; &amp; C:/Users/eleni/AppData/Local/Programs/Python/Python311/python.exe c:/Users/eleni/Desktop/bigData23/Erwtima2.py warm goat cheese toasts with honey rosemary and walnuts 5 2012-04-14 c p a burgers 5 2012-03-19 black eyed peas supper 5 2012-02-02 mashed potatoes with variations 5 2012-05-02 mashed potatoes with variations 5 2012-05-03 mashed potatoes with variations 4 2012-05-04 mashed potatoes with variations 4 2012-05-05</pre>
Εμφάνιση όλων των λεπτομερειών για την ταινία «chic greek salad» (κατηγορία δυσκολίας, διατροφικές αξίες, βήματα, περιγραφή, μέση βαθμολογία)	<pre>PS C:\Users\eleni\Desktop\bigData23&gt; &amp; C:/Users/eleni/AppData/Local/Programs/Python/Python311/python.exe c:/Users/eleni/Desktop/bigData23/Erwtima2.py 205331 4.3333333333333333 eukolh a super veggiepacked salad courtesy of rachael ray 7 chic greek salad 3389 300 350 900 430 140 70 330</pre>
Εμφάνιση των ταινιών της κατηγορίας «εύκολη» ταξινομημένες ως προς τη μέση βαθμολογία τους	<pre>strawberry daiquiri smoothie alcoholic 2.857142857142857 eukolh beef jerky cowboy style beef jerky 2.875 eukolh hopple popple german breakfast casserole 2.875 eukolh bryanna s vegan pecan pie 2.875 eukolh weight watchers baked oatmeal 2.875 eukolh perfectly candied pecans 2.875 eukolh vegan brown rice pudding 2.875 eukolh</pre>
Εμφάνιση των συνταγών που περιέχουν την ετικέτα “superbowl” με ταξινόμηση ανά	<pre>cajun shrimp cakes with andouille sausage superbowl 2007-01-16 super bowl popcorn superbowl 2007-01-15 kittencal s classic club sandwich superbowl 2007-01-15 super bowl chili superbowl 2007-01-15 italian sausage meatloaf sandwich superbowl 2007-01-15 easy ginger lover s skillet beef superbowl 2007-01-14 cheesy pork rice and black bean casserole superbowl 2007-01-14</pre>

ημερομηνία προσθήκης (πιο πρόσφατες πρώτα)	
Εμφάνιση των 20 συνταγών με την υψηλότερη μέση βαθμολογία για την ετικέτα "cocktail".	

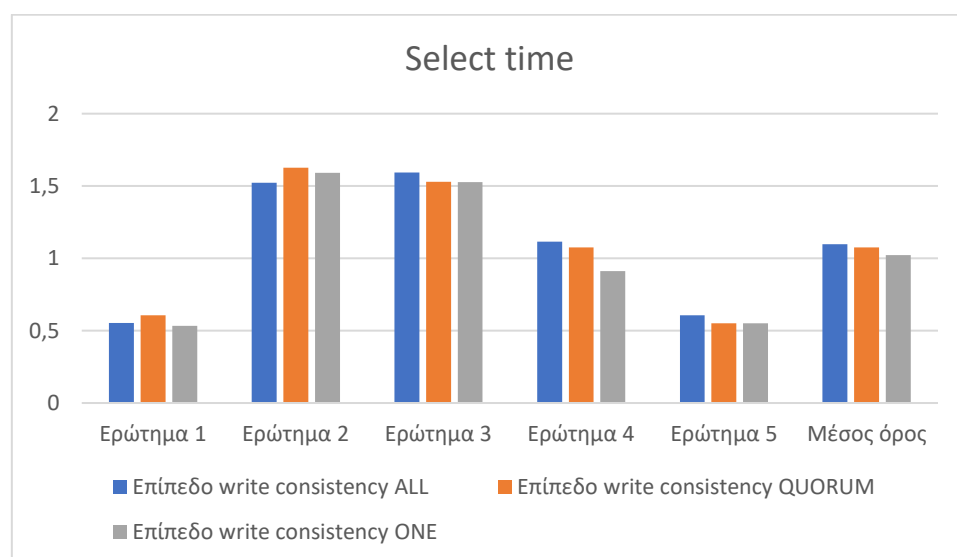
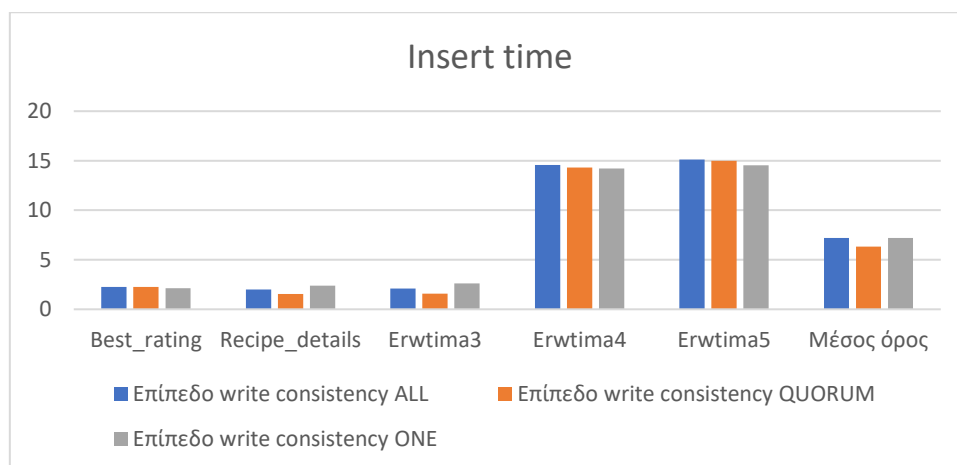
#### Ερώτημα 4A: Χρόνοι εισαγωγής δεδομένων

	Επίπεδο write consistency		
	ALL	QUORUM	ONE
<b>Best_rating</b>	2min 26sec	2min 27sec	2min 13sec
<b>Recipe_details</b>	2min	1min 54sec	2min 4sec
<b>Erwtima3</b>	2min 10sec	1min 58sec	2min 6sec
<b>Erwtima4</b>	14min 56sec	14min 32sec	14min 20sec
<b>Erwtima5</b>	15min 11sec	15min	14min 54sec
<b>Μέσος όρος</b>	7min 20sec	6min 34sec	7min 19sec

#### Ερώτημα 4B: Χρόνοι ανάκτησης δεδομένων

	Επίπεδο write consistency		
	ALL	QUORUM	ONE
<b>Select 1</b>	0.553393	0.605428	0.53241
<b>Select 2</b>	1.622467	1.615704	1.59102
<b>Select 3</b>	1.592364	1.530023	1.52713
<b>Select 4</b>	1.114568	1.075834	0.91139
<b>Select 5</b>	0.606163	0.551754	0.550225
<b>Μέσος όρος</b>	1.097791	1.0757476	1.022471

## Ερώτημα 4Γ: Σχολιασμός αποτελεσμάτων



Όπως φαίνεται στο διάγραμμα, ο χρόνος εκτέλεσης με consistency level ALL είναι ο πιο χρονοβόρος, ενώ με ONE τα λιγότερο. Αυτό συμβαίνει επειδή με το ONE στοχεύουμε σε έναν μόνο συγκεκριμένο κόμβο για την ανάκτηση των δεδομένων και δεν απαιτείται να αναζητήσουμε τα δεδομένα σε όλους τους κόμβους του δικτύου.

## Βιβλιογραφία

[1] "DataStax Bulk Loader Overview," GitHub, Sep. 21, 2023. <https://github.com/datastax/dsbulk> (accessed Sep. 24, 2023).

[2] Draw.io, "Flowchart Maker & Online Diagram Software," [app.diagrams.net](https://app.diagrams.net/), 2023. <https://app.diagrams.net/>



[3] “DataStax managed services powered by Apache Cassandra | DataStax Astra DB,” DataStax. <https://www.datastax.com/products/datastax-astra>

[4] Διαφάνειες του μαθήματος