

ΒΑΣΕΙΣ ΔΕΔΟΜΕΝΩΝ

Ομάδα: Κωνσταντίνος Κασφίκης – 2013030108

Στο 2^ο μέρος της εργαστηριακής άσκησης ζητήθηκε η δημιουργία συνάρτησης – συνδιαλλαγής και περιορισμών ως προς το πλήθος των μηνυμάτων, αιτήσεων συστάσεων και συνδέσεων.

Ο περιορισμός αυτός εφαρμόστηκε με την χρήση triggers τα οποία είναι τα εξής:

- Limit messages: Περιορίζει το πλήθος των εισερχόμενων μηνυμάτων κάθε χρήστη στα δέκα ανά ημέρα.
- Limit invitations: Περιορίζει το πλήθος των εισερχόμενων αιτήσεων σύνδεσης κάθε χρήστη σε πέντε ανά ημέρα.
- Limit rec request: Περιορίζει το πλήθος των εισερχόμενων αιτήσεων σύστασης κάθε χρήστη σε πέντε ανά ημέρα.

Συγκεκριμένα τα παραπάνω γίνονται θέτωντας before insert triggers και έπειτα μετρώντας με την εντολή Select Count το πλήθος των εγγραφών στον εκάστοτε πίνακα που αφορούν τον χρήστη για τον οποίον δημιουργείται η νέα εγγραφή.

Το ζητούμενο Transaction δημιουργήθηκε με την μορφή συνάρτησης τύπου που επιστρέφει Void και σκοπός του είναι είτε η εκπλήρωση του συνόλου των εντολών που υπάρχουν στην συνάρτηση αυτή, ή εάν προκύψει κάποιο σφάλμα ή δεν πληρούνται οι περιορισμοί η αποτροπή εκτέλεσης ολόκληρης της συνάρτησης με χρήση της εντολής Rollback.

Τα παραπάνω εκτελούνται με την κλήση της συνάρτησης Bulkactions.

Στη συνέχεια ζητήθηκε η βελτιστοποίηση των ερωτήσεων ανάκτησης από την 1^η εργαστηριακή άσκηση.

Συνολικά οι πίνακες οι οποίοι χρησιμοποιούνται σε ερωτήσεις ανάκτησης είναι οι παρακάτω:

- EDUCATION: Χρησιμοποιούνται τα πεδία School, ToYear, FromYear, Email
- MEMBERS: Χρησιμοποιείται το πεδίο Email
- ARTICLE: Χρησιμοποιούνται τα πεδία: Email, ArticleID
- ARTICLE_COMMENT Χρησιμοποιούνται τα πεδία: Email, ArticleID
- ADVERTISEMENT Χρησιμοποιούνται τα πεδία: JobType, Salary, Industry, Country, AdvertisementID
- JOB_SEEK: Χρησιμοποιείται το πεδίο ArticleID
- JOB_OFFER: Χρησιμοποιείται το πεδίο ArticleID

Επίσης για την υλοποίηση των ερωτημάτων ανάκτησης 1,2 και 8 χρησιμοποιείται προσωρινή όψη που ονομάζεται MyView η οποία περιέχει το σύνολο των συνδέσεων μεταξύ των χρηστών. Επειδή η όψη δεν είναι materialized δεν μπορούν να εφαρμοστούν indexes. Ακόμη ευρετήρια δεν εφαρμόστηκαν στο Email από τον πίνακα MEMBER, στο ArticleID από τον ARTICLE, JOB_OFFER, JOB_SEEK καθώς αποτελούν primary keys και εξ ορισμού διαθέτουν ήδη.

Όσον αφορά τον πίνακα EDUCATION, δημιουργήθηκαν indexes τύπου BTree στα πεδία ToYear, From Year, καθώς χρησιμοποιούνται για την εύρεση εύρους τιμών,

και Email (το οποίο κανονικά θα έπρεπε να είναι τύπου Hash εφόσον χρησιμοποιείται μόνο για ισότητα, αλλά δεν υποστηρίζεται το Clustering σε Indexes τέτοιου τύπου) και index τύπου Hash στο πεδίο School καθώς χρησιμοποιείται μόνο ως κριτήριο ισότητας. Εφαρμόστηκε και Cluster ως προς το index του πεδίου Email

Στον MEMBERS δεν εφαρμόστηκε κάποιο index καθώς το Email αποτελεί Primary key.

Στον ARTICLE δημιουργήθηκε index στο πεδίο Email τύπου BTree (το οποίο κανονικά θα έπρεπε να είναι τύπου Hash εφόσον χρησιμοποιείται μόνο για ισότητα, αλλά δεν υποστηρίζεται το Clustering σε Indexes τέτοιου τύπου) και Cluster ως προς το index του πεδίου Email.

Στον ARTICLE_COMMENT δημιουργήθηκε index στο πεδίο Email τύπου Hash και index στο πεδίο ArticleID τύπου BTree (το οποίο κανονικά θα έπρεπε να είναι τύπου Hash εφόσον χρησιμοποιείται μόνο για ισότητα, αλλά δεν υποστηρίζεται το Clustering σε Indexes τέτοιου τύπου) και Cluster ως προς το index του πεδίου ArticleID.

Στον ADVERTISEMENT δημιουργήθηκε index στο πεδίο JobType τύπου BTree (το οποίο κανονικά θα έπρεπε να είναι τύπου Hash εφόσον χρησιμοποιείται μόνο για ισότητα, αλλά δεν υποστηρίζεται το Clustering σε Indexes τέτοιου τύπου) και index στα πεδία Country και Industry τύπου Hash και τύπου BTree στο Salary και Cluster ως προς το index του πεδίου JobType.

Στους πίνακες JOB_OFFER και JOB_SEEK δεν εφαρμόστηκαν indexes καθώς τα ArticleID αποτελούν Primary keys.

Στον παρακάτω πίνακα εμφανίζονται οι μετρήσεις των χρόνων σχεδίασης και εκτέλεσης των ερωτημάτων, πριν και μετά στην εφαρμογή indexes και clustering (M.O.) καθώς και τα indexes και cluster που εφαρμόστηκαν και αφορούν το ερώτημα αλλά και τα συμπεράσματα που προκύπτουν από την επεξεργασία των τιμών.

Ερώτ. Ανάκ.	Χωρίς Indexes/Clustering		Με Indexes/Clustering		Συμπεράσματα
	Χρόνος σχεδίασης	Χρόνος εκτέλεσης	Χρόνος σχεδίασης	Χρόνος εκτέλεσης	
1	1,726	0,932	0,817	0,965	EDUCATION.Email (BTree) Cluster EDUCATION.ToYear (BTree) EDUCATION.FromYear (BTree) EDUCATION.School (Hash) Παρατηρείται διαφορά στους χρόνους πριν και μετά την εφαρμογή ενώ με την χρήση της εντολής Explain Analyse διαπιστώνεται ότι όντως γίνεται χρήση των Indexes που δημιουργήθηκαν.
3	0,168	0,356	0,194	0,342	ARTICLE.Email (BTree) Cluster Δεν παρατηρείται σημαντική διαφορά στους χρόνους πριν και μετά την εφαρμογή ενώ με την χρήση της εντολής Explain Analyse διαπιστώνεται ότι δεν γίνεται χρήση των Indexes που δημιουργήθηκαν. Ωστόσο δεν κρίνεται απαραίτητη η αλλαγή της εντολής.
4	0,376	3,716	0,367	3,116	ARTICLE_COMMENT.Email (Hash) ARTICLE_COMMENT.ArticleID (BTree) Cluster Παρατηρείται διαφορά στους χρόνους πριν και μετά την εφαρμογή ενώ με την χρήση της εντολής Explain Analyse διαπιστώνεται ότι όντως γίνεται χρήση του Index ArticleID αλλά όχι του Email.
5	0,189	0,731	0,183	0,669	ARTICLE_COMMENT.ArticleID (BTree) Cluster Δεν παρατηρείται σημαντική διαφορά στους χρόνους πριν και μετά την εφαρμογή ενώ με την χρήση της εντολής Explain Analyse διαπιστώνεται ότι δεν γίνεται χρήση των Indexes που δημιουργήθηκαν. Ωστόσο δεν κρίνεται απαραίτητη η αλλαγή της εντολής.
6	1,564	1,180	0,608	1,209	EDUCATION.Email (BTree) Cluster ARTICLE.Email (Hash) Παρατηρείται διαφορά στους χρόνους πριν και μετά την εφαρμογή ενώ με την χρήση της εντολής Explain Analyse διαπιστώνεται ότι όντως γίνεται χρήση των Indexes που δημιουργήθηκαν.
7	0,390	3,193	0,943	1,611	ADVERTISEMENT.JobType (BTree) Cluster ADVERTISEMENT.Industry (Hash) ADVERTISEMENT.Country (Hash) ADVERTISEMENT.Salary (BTree) Παρατηρείται διαφορά στους χρόνους πριν και μετά την εφαρμογή ενώ με την χρήση της εντολής Explain Analyse διαπιστώνεται ότι όντως γίνεται χρήση των Indexes που δημιουργήθηκαν.

Γενικά παρατηρείται ότι στην εντολή GroupBy καθώς και στα Aggregate Functions δεν γίνεται χρήση των indexes που έχουν εφαρμοσθεί, ενώ στα κριτήρια του Where καθώς και στα Join τα indexes όντως συμβάλλουν στην βελτιστοποίηση των ερωτημάτων.

Για την επαρκή μέτρηση στην διαφορά των χρόνων πριν και μετά την εφαρμογή indexes και clustering δημιουργήθηκε συνάρτηση με όνομα ADD_JUNK_DATA η οποία αυξάνει τις εγγραφές στους πίνακες με την παρακάτω λογική. Γίνεται εισαγωγή 500 νέων χρηστών από τους οποίους ο καθένας θα δημοσιεύσει ένα νέο άρθρο και τρία σχόλια σε άρθρα. Κάθε χρήστης επίσης θα δημοσιεύσει από μία διαφήμιση ενώ οι μισοί από αυτούς θα δημοσιεύσουν ένα Job offer και οι άλλοι μισοί ένα Job Seek. Κατ' αυτόν τον τρόπο αυξάνεται επαρκώς ο όγκος των δεδομένων στους πίνακες στους οποίους γίνονται τα ερωτήματα ανάκτησης της πρώτης εργαστηριακής άσκησης.

Στις παραπάνω μετρήσεις οι συνολικοί χρόνοι μεταβάλλονταν ακόμη και κατά την διαδοχική εκτέλεση του ίδιου ερωτήματος οπότε σαν τιμή ελήφθη ο M.O. αρκετών δοκιμών, και για τον λόγο αυτό δεν πρέπει να χρησιμοποιηθούν σε απόλυτες συγκρίσεις, ωστόσο καταδεικνύουν την βελτιστοποίηση των ερωτημάτων στα οποία έχουν εφαρμοσθεί indexes και clustering.

Τέλος υλοποιήθηκε εξωτερική εφαρμογή γραμμένη σε περιβάλλον Java με την χρήση του οδηγού για σύνδεση με τη βάση Postgres JDBC