Linnéuniversitetet
Kalmar Växjö

Linnaeus University
2DV513 - Database Theory

Assignment 2

Author: Rashed QAZIZADA
Area: Computer Science
Supervisor: Ilir Jusufi
Course code: 2DV513
Semester: Autumn 2020

Abstract

This page was intentionally left unchanged.

Keywords

Place your keywords here

Contents

1 Task : Relational algebra

1.1. Π_{name} (σ code = '2dv513' (student $\bowtie$ enrolledIn))

1.2. Π_{name} (σ code = '1dv513' $\cap$ code '2dv513' (student $\bowtie$ enrolledIn))

1.3. Π_{lecturer} (σ code = '2dv610' (enrolledIn $\bowtie$ subject))

1.4. Π_{lecturer} (σ code = '1dv513' $\cap$ code '2dv513' (enrolledIn $\bowtie$ subject))

1.5. Π_{names} (σ lecturer! = 'Ilir' (student X subject))

2 Task : FDs and Normalization

2.1 Finding functional dependencies.

I can assume that this database will check every week. Therefore each week is a unique.
Day time $\longmapsto$ applicant manager room.
Applicant day $\longmapsto$ time room manager.

2.2 Finding the keys of the relation

Day and time are to be the keys of the relation.

2.3 Showing that the relation is in 3NF but not in BCNF

The relation should not have partial dependency neither transitive dependency. This means the relation will be in 3NF if it satisfies 2NF.

- Checking for partial dependency:

  Since we have day and time as the composite keys. Therefore, I cannot know who is the applicant only from knowing the day value or the time value. Furthermore, the same rule applies to the manager, and still, I cannot know who is having the time or the day. Therefore, there is no partial dependency. So, it is in 2NF.

- Now, let's check for transitive dependency:

  From the transitive dependency perspective, I could not determine who is the applicant only from knowing the manager or the room. Therefore, there is no transitive dependency. This means the relation has 3NF normalization.

- Now, showing that it does not satisfy BCNF.

Since, we have a non-prime attribute manager which determines a prime attribute which is the time then this relation is not in BCNF. For this I assume that the Manager can have just one meeting for each day.

2.4 Decomposing the relation in relations that are in BCNF.

In order to compose the relation to make relations in BCNF, the primary key of the relations needs to be changed. So, the relations will be as following:
People (M_ID,Manager, applicant)
date (day, time, M_ID, room)

Now we can see that the above relations satisfy 3NF normalization, and it does not have a non-prime attribute which determines a prime attribute, Therefore, it satisfies BCNF as well.

2.5 E/R diagram that describes the system with incorporation of all dependencies

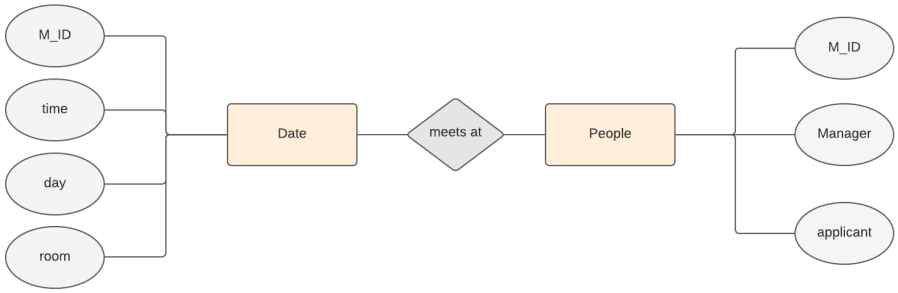


Figure 1: E/R diagram

3 Task : Setting up the Reddit database

3.1 An E/R diagram for my design as well as schemeas with their types

The E/R diagram shows the relations between a reddit comment with a subreddit and links.

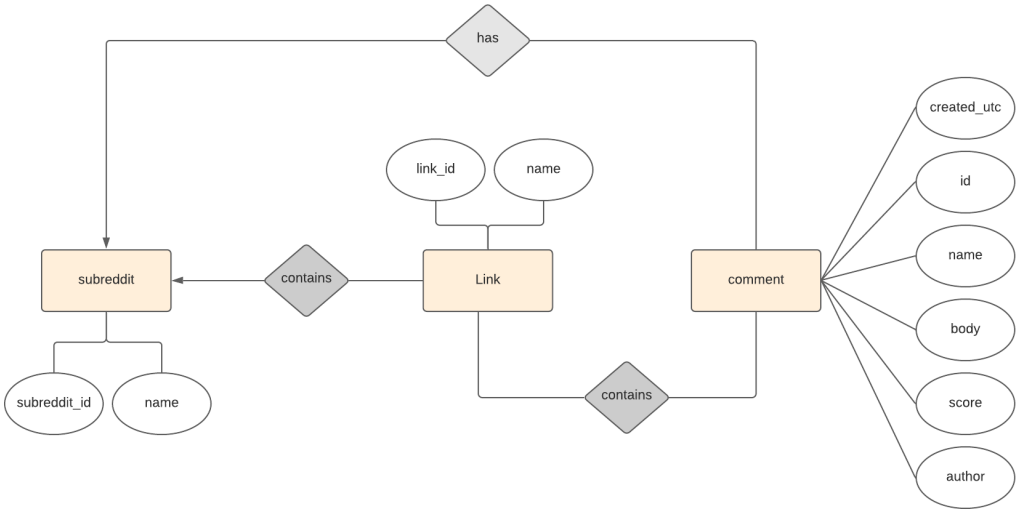


Figure 2: E/R diagram with their types

Schemas with types:

Comment: has subreddit_id and link_id as foreign keys

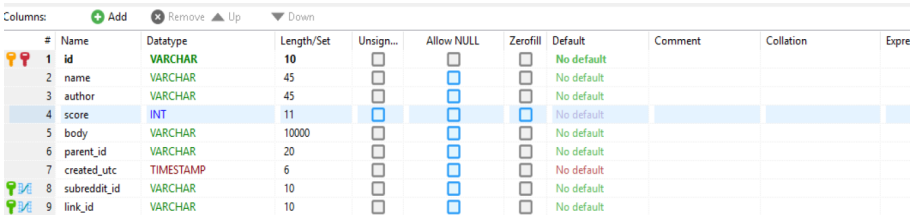


Figure 3: subreddit_id and link_id

Link: has subreddit_id as foreign keys.

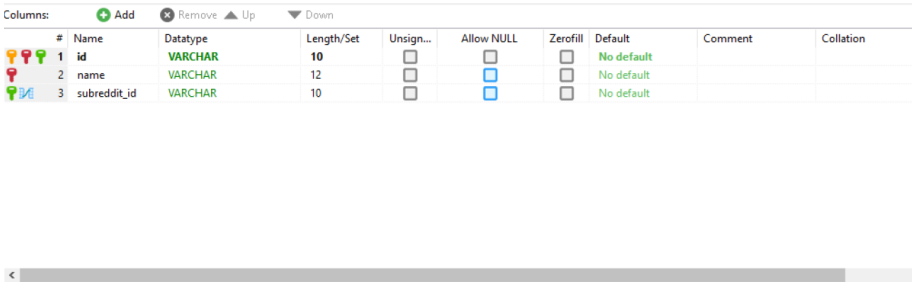
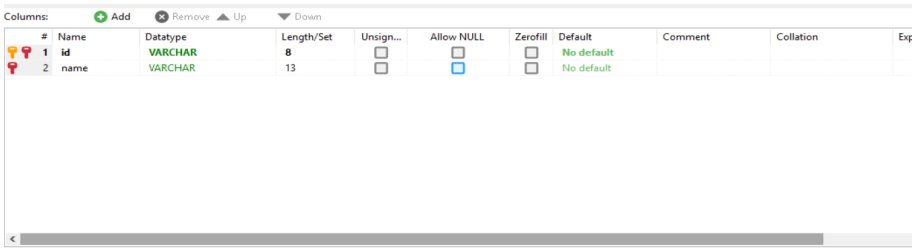


Figure 4: subreddit_id

Subreddit:



4 Task: Importing data

- First, I populated tables with constraints the process almost ended in 29 minutes just to migrate the entire data to the database.

```
150428 errorrrrrr Subreddit ER_DUP_ENTRY Duplicate entry 't5_2cneq' for key 'PRIMARY'
150428 errorrrrrr Link ER_DUP_ENTRY Duplicate entry '5zep2' for key 'PRIMARY'
150428 added Comment
150429 errorrrrrr Subreddit ER_DUP_ENTRY Duplicate entry 't5_6' for key 'PRIMARY'
150429 errorrrrrr Link ER_DUP_ENTRY Duplicate entry '5zik0' for key 'PRIMARY'
150429 added Comment
```

- Second, I tested without constraints which means no primary keys nor unique or foreign keys were assigned and the process took almost 22 minutes just to get finished. Furthermore, the table was full of duplicates data.

4.1 Conclusion

Comparing the results of populating the data with and without constraints, I can see that the insertion without constraints is faster. However, to migrate the data without constrains I achieved faster results on the other hand we should also consider the duplication of the data, Furthermore, we should avoid adding constraints after importing it or I could choose some rows even if I have duplication with data just by using some specific keyword. However, it is a good approach to use constraints before populating or migrating the data to the database.

5 Task: Queries

How many comments have a specific user posted?
In order to get the number of comments for a specific user and whereas the author is a specific username then I could achieve this just by SELECT COUNT(ID) FROM comment WHERE AUTHOR =’igiveyoumylife’ the result of this specific user that posted is 38.

```
1 SELECT COUNT(ID) FROM comment WHERE AUTHOR = 'igiveyoumylife'
```

Result #1 (1r × 1c)

COUNT(ID)
38

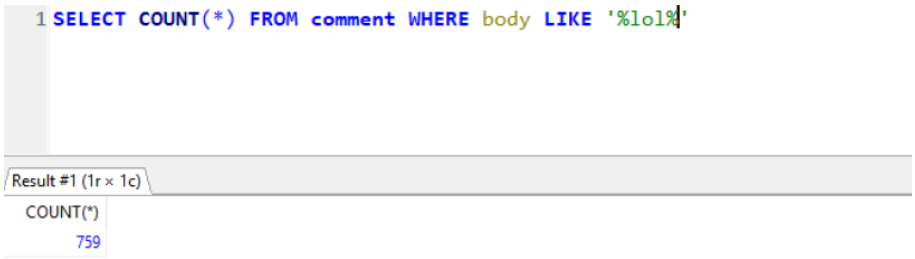
How many comments does a specific subreddit get per day?
I selected the “bugs” subreddit for each day, in total a specific subreddit get 104 comments per day.

```
1 select COUNT(Id) from comment WHERE subreddit_id = 't5_2h84o' GROUP BY created_utc
```

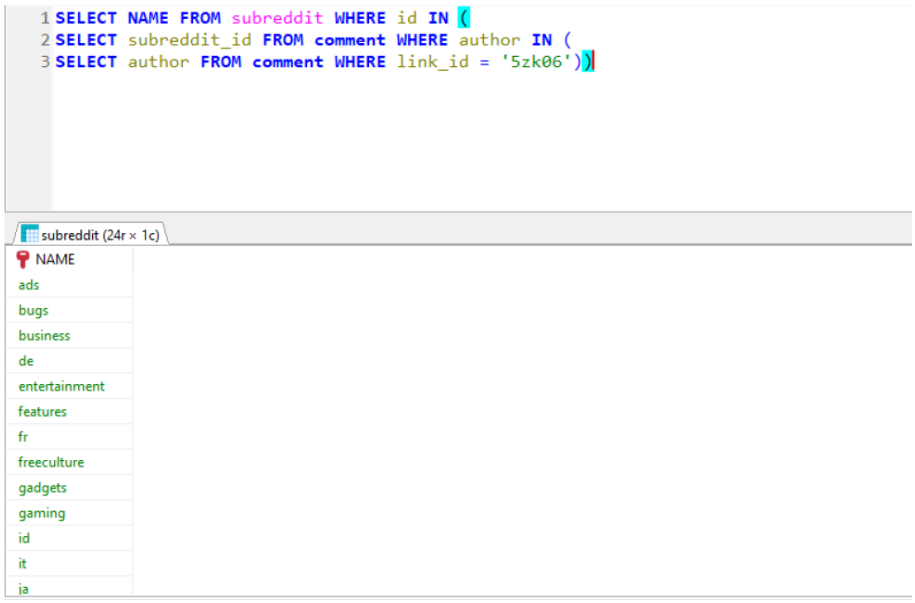
Result #1 (104r × 1c)

COUNT(Id)
1
1
1
1
1
1

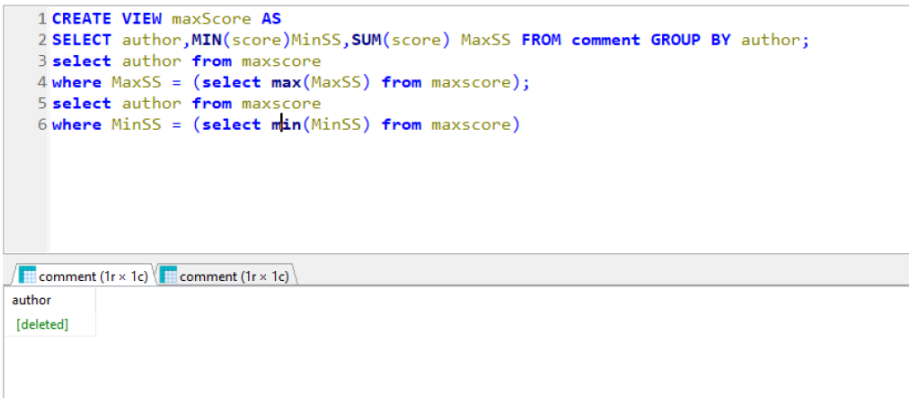
How many comments include the word 'lol'?
In this case "SELECT COUNT(*) FROM comments WHERE body LIKE '%lol%' " I am counting the comments where the string 'lol' is contained in that string. The percent signs before and after will ensure that any string that contains lol in it counts it like as following: lollipop, lolypop, loblolly etc.



Users that commented on a specific link has also posted to which subreddits?
I am using the link to get the author name then I selected all subreddit_id from the same author.. In this case the query will show all the data that fulfils my condition.



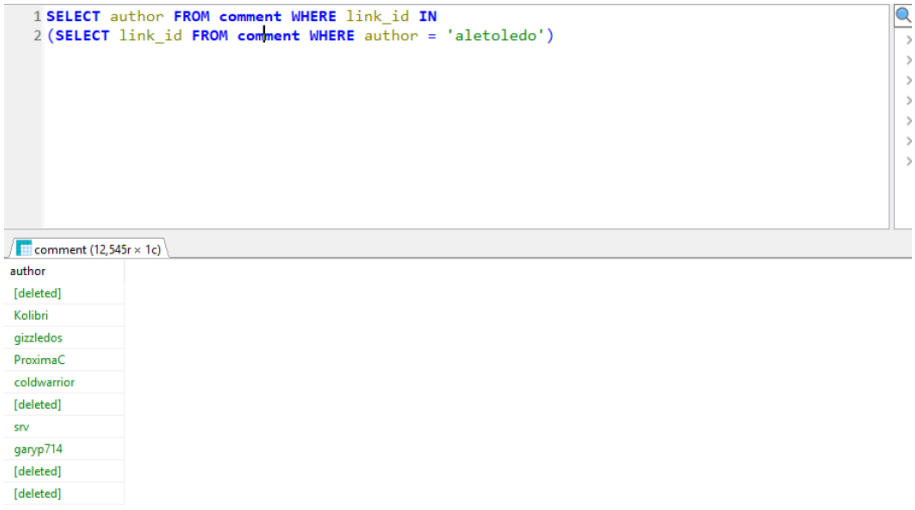
Which users have the highest and lowest combined scores? (combined as the sum of all scores)
CREATE VIEW maxScore AS SELECT author, MIN(score)MinSS,SUM(score) MaxSS FROM comment GROUP BY author; select author from maxscore where MaxSS = (select max(MaxSS) from maxscore); select author from maxscore where MinSS = (select min(MinSS) from maxscore)



Given a specific user, list all the users he or she has potentially interacted with (i.e., everyone who as commented on a link that the specific user has commented on)

In this case duplicated authors are interacted. I had to be careful not to duplicate the authors.

SELECT author FROM comment WHERE link_id IN
(SELECT link_id FROM comment WHERE author = 'aletoledo';



Studen id: rq222ah

THE END

Which users has only posted to a single subreddit?
Here, I need to select all the authors who commented exactly one subreddit.

SELECT COUNT (subreddit_id) AS SubAmount, author
FROM comment GROUP BY author HAVING COUNT(SubAmount) =1