PSG College of Technology - Coimbatore
Department of Applied Mathematics and Computational Sciences
IV M.Sc. Theoretical Computer Science – RDBMS Lab
Worksheet – 4

This dataset is the blog posts of the Techcrunch technology blog which was collected for a research paper "Identifying the Productive and Influential Bloggers in a Community" to identify members who are both productive and influential. (https://www.kaggle.com/lakritidis/identifying-influential-bloggers-techcrunch?select=inlinks.csv)

This database consists of four files. Precise descriptions and record counts for each are provided below.

1. **Authors.csv** : A list of the bloggers of Techcrunch, along with their (unique) IDs and some statistics. This file includes the records of the 107 bloggers of Techcrunch. Columns 3 and 4 contain the computed values for MEIBI and MEIBIX, two popular blogger evaluation metrics.

   Author(Author_ID, author_name, MEIBI metric, MEIBIX metric)

- **posts.csv**: A database of the retrieved blog posts, It includes 19464 posts authored by 107 bloggers. MEIBI and MEIBIX are two well-established methods for evaluating blog posts and bloggers (see the last two columns). Additional statistics are also recorded, such as the number of words in the post (with and without stop words)

Posts(

post id,	Authorid	post title	post content ,	post URL,	Date,	number of words,	number of words without stop words,	MEIBI metric,	MEIBIX Metric

)

comments.csv: This file contains the 746561 comments which were submitted by the readers of the 19464 posts of Techcrunch**.**

Comments (Comment_ID, post_Id, comment_content,Date)

inlinks.csv: it contains 193808 pages that contain links to the 19464 posts of Techcrunch.

Inlinks(inlink_ID, post_ID,inlink_title, date, inlink_URL)

Access link:

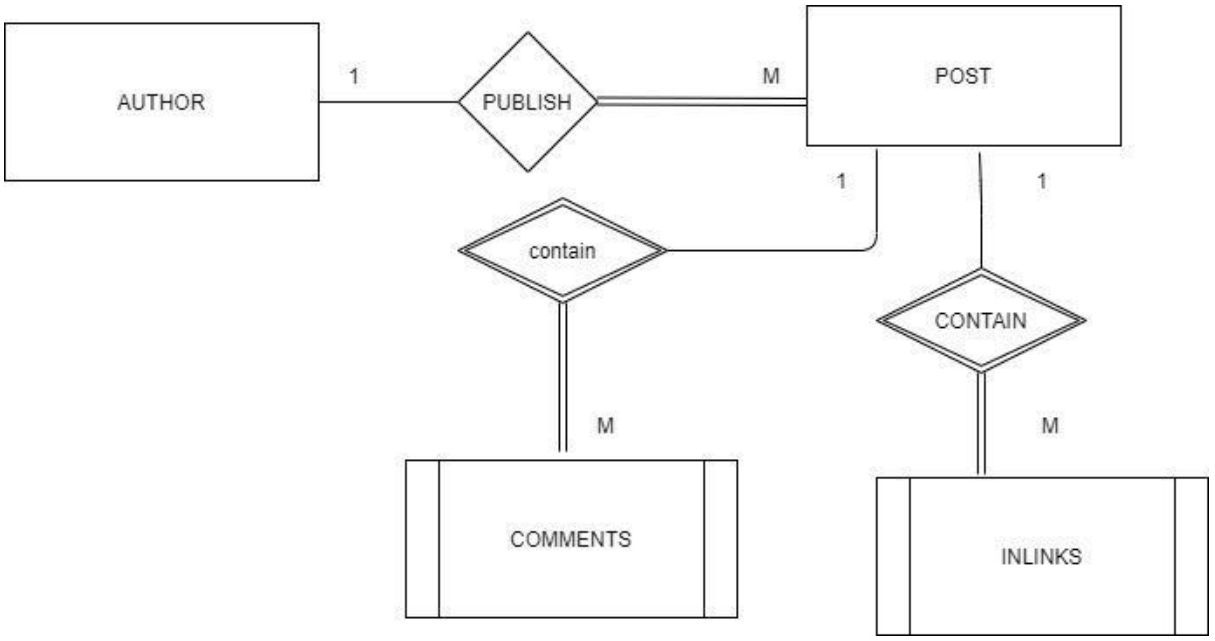
https://drive.google.com/file/d/1buhOSP3H82crPzwIttPt_5pmL4mHEp8v/view?usp=sharing

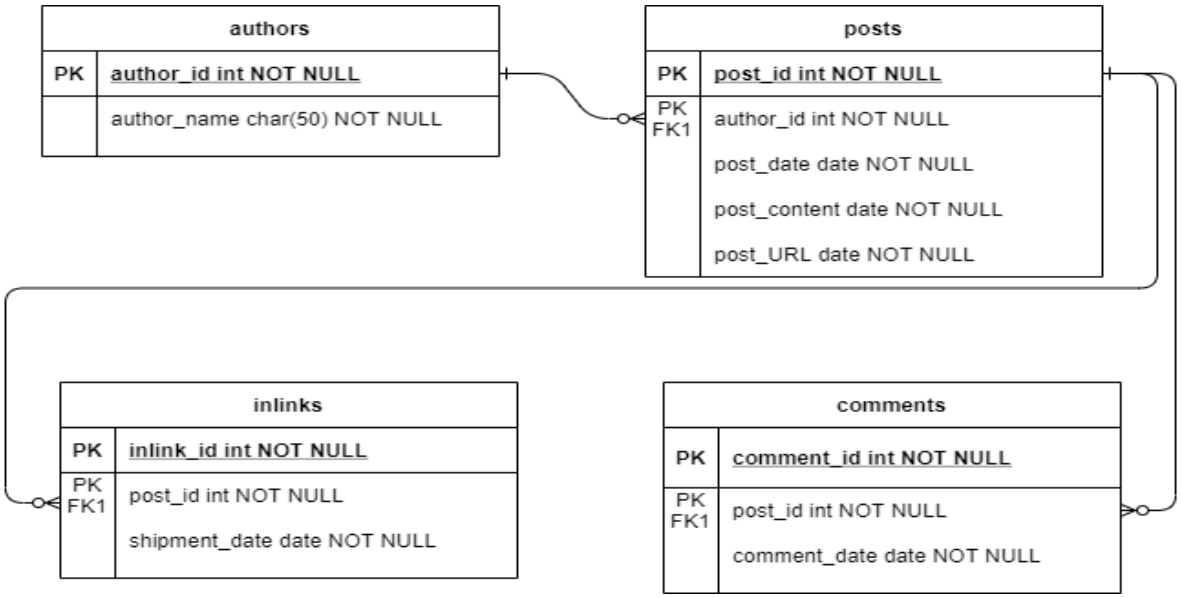
You can access the above tables from the user C3328 table space
eg: select * from c3328.AUTHORS;

ı] **Write DDL statements to create key constraints**

2) Write a DDL to set a sequence on inlink_id for generating ids that begins with 1, increments by 1. Analyse your problem statement and add the necessary strategies for foreign keys. Eg:. ON DELETE/ON UPDATE CASCADE/NULL/DEFAULT.

The given diagram visualizes the Conceptual schema.(not all attributes are shown) and its equivalent Relational DB schema





II) Write **interactive** PL-SQL Block/Stored Procedures/Functions to perform the following. You should use only JOINS for retrieval operations. Try to avoid using subqueries. Use implicit and explicit cursors wherever necessary.

Without PL/SQL, Oracle9*i* must process SQL statements one at a time. Each SQL statement results in another call to the Server and higher overhead. However, with PL/SQL, you can send an entire block of SQL statements to the server. This minimizes communication between your application and the server. Thus PL/SQL can help you reduce overhead, improve performance, and increase productivity.

Anonymous Pl/sql Blocks with subprograms such as procedures and functions

a. List author names along with number of posts authored by them

b. Get the number of posts written by 'Mr.**Marc Benioff'.** Using implicit cursor

c. Display post_title, post_date and URL published by author ' John Biggs'. Arrange the result in the order of date published

d. Find the number of comments written for each post. Display post title, date, URL and number of comments, Order result based on number of comments

e. Find the author name(s) whose posts had received the highest number of comments

Stored Procedures

a. Fetch the details such as title, date, url published in 2020 from POST. Process the posts to Rank them in terms of the number of words. Posts with above 500 words are ranked as A, between 200 to 500 as B , below 200 as c and others F.

a. Print author names, post title, date and URL for which more than 2 comments (get the no. of comments as input) had been written

b. write code to Increase the MEIBI index in the author table by 2% if at least one of the posts contains more than 2000 words. Get the number records updated.

c. For each author, list his/her name, number of posts authored, total comments written and number of inlinks. Your cursor pulls posts which contain above 1000 words.

d. Filter the details of authors who wrote more than 200 words in all of his/her posts. Further add code to list the number of posts with above 1000 words , 500 to 1000 words and below 500 words for each author.

e. Find the names of authors who received at least one comment for each of his/her posts (division)

f. Fetch the top 10 influential authors using explicit cursor, For each author, compute the number of posts, the total and average number of words used

g. Find authors who have written more posts than 'Jason Kincaid'(INPUT) Print the author name and number of posts. Use implicit cursor for obtaining the number of posts for each author

h. Get the authors names of all the posts with comments and without comments (full outer join)

i. Find the authors who authored more than Mike Butcher(input)

Create Functions for the following tasks

1. Identify the most influential and non_influential author based on the MEIBIx metric.
2. Compute the average MEIBI metric score for the given author . Accept author name as input. (sum(MEIBI)/total posts for an author)
3. Find the post_title which has only comments but no inlinks (set minus)
4. On which day, the most posts had been authored ?
5. In 2010, how many posts have been published?
6. Which author's posts got the highest number of inlinks?

7. On which date, the highest number of posts have been published in 2010?