

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](https://www.kaggle.com/lakritidis/identifying-influential-bloggers-techcrunch?select=inlinks.csv)" 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(

<u>post_id</u>	Authorid	post title	post content	post URL	Date	number words	of	number words	of	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:

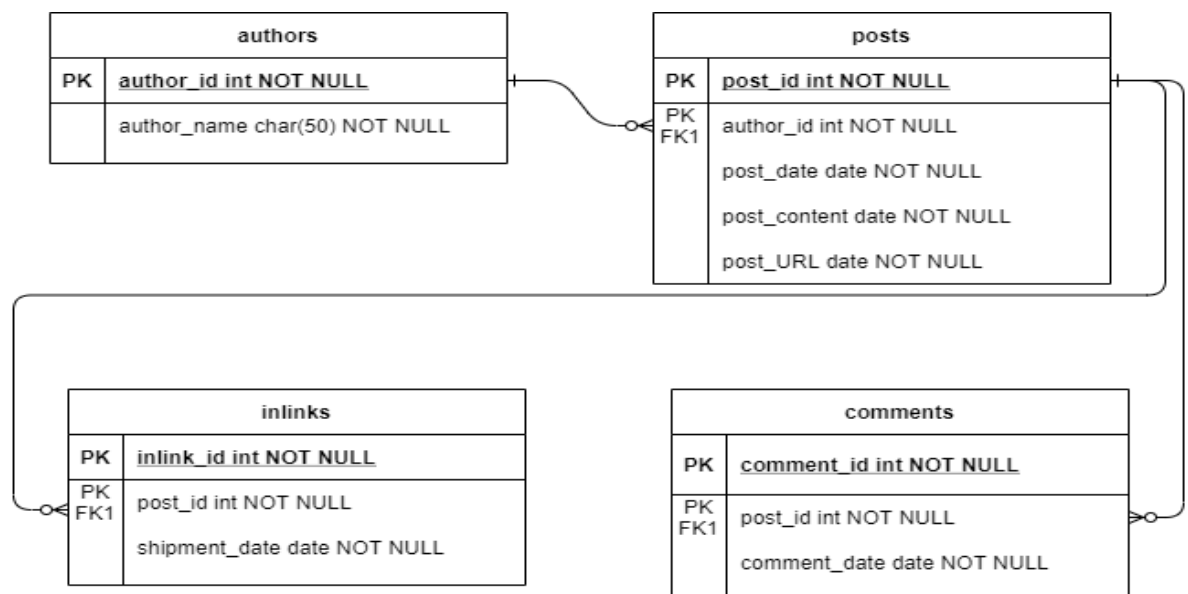
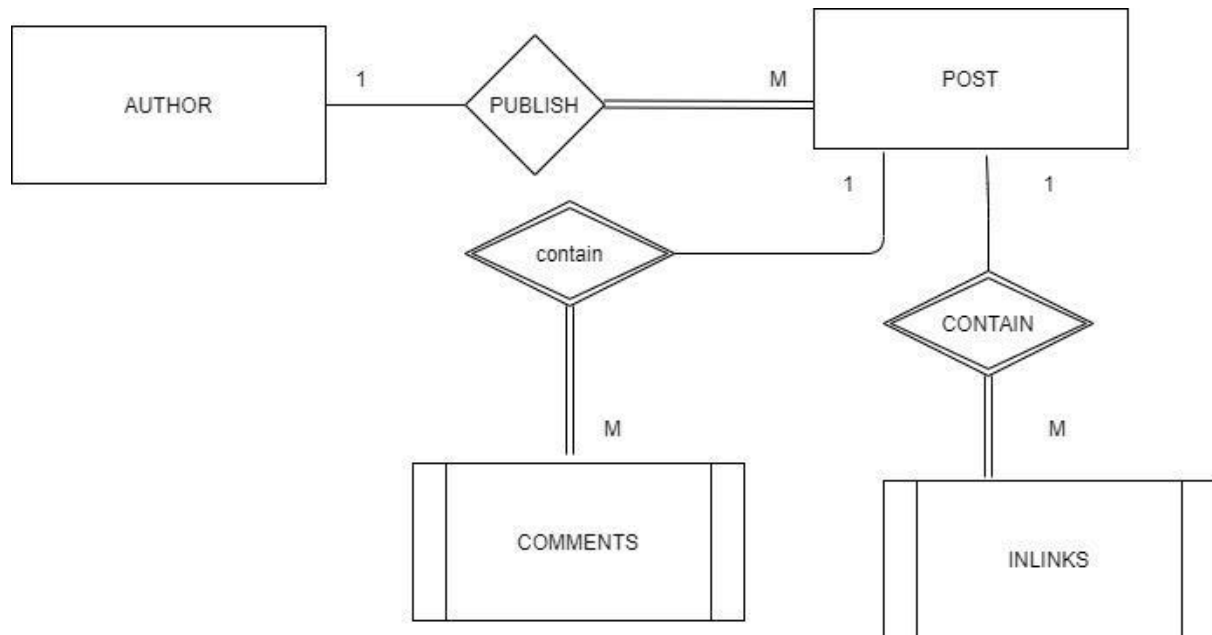
https://drive.google.com/file/d/1buhOSP3H82crPzwIttPt_5pmL4mHEp8v/v?usp=sharing

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

η Write DDL statements to create 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.

Pl/sql Blocks and procedures

- List author names along with number of posts authored by them
- Find all information of authors who have published post on '25/03/2010'
- List the authors who have not authored any post.
- Get the number of posts written by 'Mr.**Marc Benioff**'
- Display post_title, post_date and URL published by author ' John Biggs'.
Arrange the result in the order of date published
- On which date, the highest number of posts have been published in 2010?
- Get the names of authors who wrote more than 3 posts
- 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

- i. Find the author name(s) whose post had received the highest number of comments
- j. Find the authors who authored more posts than Mike Butcher

Stored Procedures

- k. Find authors who received comments for at least one post.
- l. For each post, list details such as title, date, url and number of comments. Print both posts with comments and posts with no comments.
- m. List those authors who have published posts only on Sunday.
- n. Print author names, post title, date and URL for which more than 2 comments (no. of comments - input) had been written
- o. Get the post title, date of posts that have more than 3 inlinks.
- p. Find authors whose posts had no inlinks and no comments.
- q. For each author, list his/her name, number of posts authored, total comments written and number of inlinks.
- r. Find the authors who wrote more than 200 words(number of words as input) in all of his/her posts.
- s. Find the names of authors who received at least one comment for each of his/her posts (division)
- t. Find the ids and names of authors who have authored two different posts on the same day.
- u. For each author, compute the number of posts, the total and average number of words used in all the posts.
- v. Find authors who have written more number of posts than 'Jason Kincaid'(IN)
- w. Get the authors names of all the posts with comments and without comments (full outer join)
 - x. Write a query to obtain the following result

Post_title,	Date,	number of inlinks
Twitter Tweaks Search, A Few Popular Tweets Now At The Top Of Results	01-04-2010	1
-----some---	-----	null

Create Functions for the following tasks

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

7. Which author's posts got the highest number of inlinks?