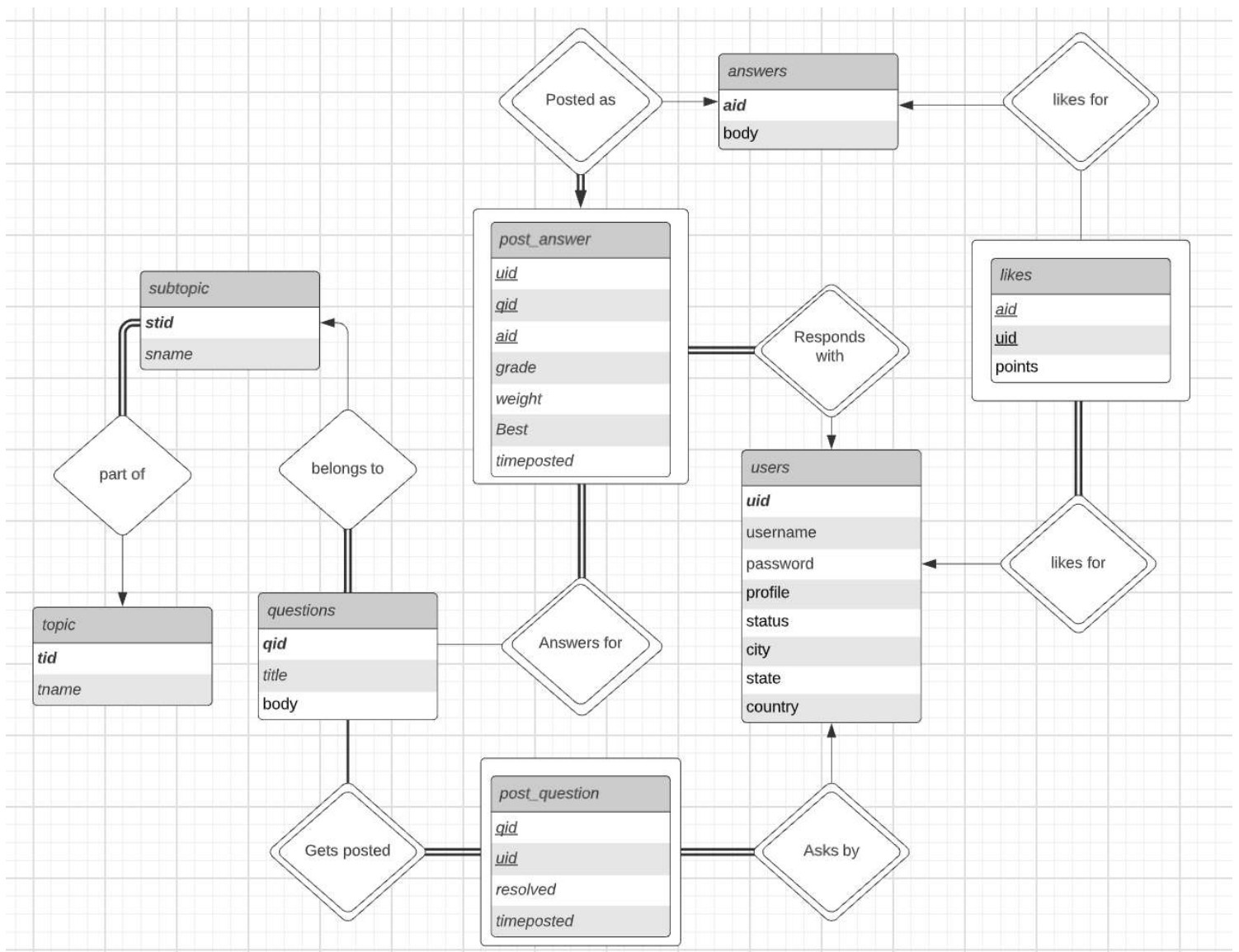


Databases Project Part 1

(a) Design, justify, and create an appropriate relational schema for the above scenario. Make sure your schema avoids redundancies. Show an E-R diagram of your design, and translate it into a relational format. Identify keys and foreign key constraints. Note that you may have to revisit your design if it turns out later that the design is not suitable.

E-R Diagram:



- Topic hierarchy is defined as a two-level system where the high-level topics are placed in the topic table
 - For each topic in the topic table, there can be a corresponding set of attributes in the subtopic table where a subtopic can be created as a subcategory of one of the main topics found in the topic table.
 - For a main topic of “Computer Science,” the subtopics in the subtopic table may include Robotics, Programming, and Software Development which will be located in the subtopic table as part of the lower hierarchy of Computer Science. These attributes will use topic’s tid to indicate that these subtopics belong to one main topic.
- For this schema, we’re assuming that a user can assign a question to only one subtopic (and topic by proxy) because it will be assumed that our subtopics will not overlap in terms of content.
 - Topics and subtopics are not weak constraints due to these tables being given a predefined list of topics and subtopics to choose from. These options will exist on their own and are not created by a user.
 - However, every subtopic must be related to a main topic.
- If a question is not related to any of the topics provided, then it will be marked as “other” and “off-topic” for the topic and subtopic category.
- Answers do not have a topic because every answer must be tied to a question. Therefore, the topic of every answer will be the topic of the question being answered.
- While post_answer’s grade attribute increases whenever a point from the likes table is given, the grade and weight will be modified if an answer is marked as best answer, so these two tables must be independent even if they provide some redundancy.

Implementation:

1. When a user signs up, he/she will be inserted into the users table to be identified as a new user.
2. If the user wants to ask a question, he/she will post the question to the question board where the question will be inserted into the questions table with the provided title for the question and body where the main question will be asked.
 - a. The user will also have to include one subtopic corresponding to the category of the question being asked.
 - b. These subtopics will be provided by a predefined list of subtopics to choose from. These subtopics were created as part of a more general main-topic (Databases is a subtopic that belongs to the main topic of Computer Science).
3. Once the question is posted, the data regarding the question’s post will be recorded and inserted into post_questions.
 - a. Post_questions will store the id of the question, the id of the user who posted the question, timestamp of the moment the question was posted, and a flag if the question was resolved.
 - b. The question will be marked as unresolved by default.
4. If a user sees a question and wishes to place an answer, he/she will write an answer that will be inserted into the answers table.

- a. The answers table will record the answer id and the message body.
 - b. Because the answer is directly tied to its target question, some of the traits of the question's traits will be inherited by the sets of answers as well (such as same topic and subtopic).
5. The answer is tied to the user who posted the answer and the question being answered in the post_answer table.
 - a. This table will assign a weight to the question to determine its value. This value will be given based off the user's status (beginner, intermediate, or advanced)
 - b. The grade is the ratings provided to it by other users through likes.
 - c. The time when the answer was posted is also recorded as a timestamp.
6. When the answer is posted, other users may provide a rating by giving the answer a like. This will add a like by inserting into a likes table where the user who liked it would be recorded along with the answer.
 - a. This will give the answer a point which will increase the answer's grade in post_answer.
7. If the person who asked the question sees a valuable answer, he/she can mark their answer as resolved, turning the resolved attribute to true.
 - a. The user may also assign the best answer. This answer will be marked in post_answer where the best boolean is turned into true and is given additional weight and grading.

(b) Use a database system to create the database schema, together with key, foreign key, and other constraints.

Table construction:

users(**uid**, username, password, profile, status, city, state, country)

uid: primary key

status: basic, intermediate, advanced

answers(**aid**, body)

aid: primary key

body: Not Null

FULLTEXT(Body)

CREATE FULLTEXT index body_index on answers(body);

questions(**qid**, stid, title, body)

qid: primary key

Stid references subtopic(stid)

Title: not null

FULLTEXT(Title, Body)

```
CREATE FULLTEXT index title_index on questions(title);
CREATE FULLTEXT index body_index on questions(body);
```

topic(**tid**, tname)

Tid: primary key

subtopics(**stid**, tid, stname)

stid: primary key

tid references topic(tid)

Weak relationships:

post_answers(uid, qid, aid, grade, weight, best, timeposted)

uid references users(uid)

Qid references questions(qid)

Aid references answers(aid)

Weight: not null

Timeposted: not null

post_question(qid, uid, resolved, timeposted)

qid references questions(qid)

uid references users(uid)

Resolved: Boolean

Timeposted: not null

likes(uid, aid)

uid references users(uid)

aid references answers(aid), post_answer(aid)

(c) Write SQL queries (or sequences of SQL queries or scripting language statements) for the following tasks. You may use suitable placeholder values in the statements.

(1) Create a new user account, together with username, email, password, city, state, country, and profile.

- For this question, we'll be inserting a new user named Pippin23 who decided his password will be password5.
- This user is from Los Angeles, California in the U.S and will be filling out his location in an abbreviated form.
- He has yet to fill out his profile, so PLACEHOLDER is being written in its place.
- He asked for and will be provided intermediate status upon creation of his account.

```
INSERT INTO users (uid, username, password, profile, status, city, state, country) VALUES (6, 'Pippin23', 'password5', 'PLACEHOLDER', 'INTERMEDIATE', 'LA', 'CA', 'US');
```

```
9 • INSERT INTO users (uid, username, password, profile, status, city, state, country) VALUES (6, 'Pippin23', 'password5', 'PLACEHOLDER', 'Basic', 'LA', 'CA', 'US');
```

uid	username	password	profile	status	city	state	country
1	JV4	password	PLACEH...	BASIC	Chicago	CO	US
2	Bobby2	password1	PLACEH...	Expert	New York City	NY	US
3	Mike6	password2	PLACEH...	BASIC	Baton Rouge	LA	US
4	Janet1	password3	PLACEH...	BASIC	Houston	TX	US
5	Daryl9	password4	PLACEH...	Advanced	Las Vegas	CA	US
6	Pippin23	password5	PLACEH...	Basic	LA	CA	US
7	Darrin4	password6	PLACEH...	BASIC	Austin	TX	US
8	Mang0	password7	PLACEH...	BASIC	Boston	NY	US
9	Leffen	password8	PLACEH...	BASIC	Seattle	WA	US
10	Elphelt	password9	PLACEH...	BASIC	Dallas	TX	US
	NULL	NULL	NULL	NULL	NULL	NULL	NULL

The highlighted row shows the newly inserted user into the table.

(2) Insert a new question into the system, by a particular user and assign it to a particular topic in the hierarchy.

- Because the questions table is split into a questions table and a post_question table, inserts for both of these tables will be shown as part of the process of adding a new question to the system.
- The user posting the question is uid = 3, username Mike6.
- The subtopic stud = 6 is labeled as "OTHER" and is linked to the main topic labeled as "Off-Topic"





INSERT INTO questions (qid, stid, title, body) **VALUES** (5, 6, 'HOW DO I FLY', 'I try flapping my arms, but it doesn't work.');

qid	stid	title	body
1	3	HOW TO 1+1	HELP, HOW DO I COMMIT 1+1
2	4	HOW TO PAY TAXES	Help, the government is inquiring about my car's extended warranty
3	2	HOW TO GROW PLANT	My crops keep dying when I feed it coca-cola. What am I doing wrong.
4	5	HOW TO HACK	Hackerman keeps demanding ransom in bittycoin. How do I pay.
5	6	HOW DO I FLY	I try flapping my arms, but it doesn't work.
6	7	Difference between ...	What's the difference between a for loop, a while loop, and a do-while...
7	3	Difference between ...	What's the difference between adding 1+1+1 and 1*3?
8	6	Kingdoms in microbio...	What are the 3 domains of life that live in the microbiology world?
9	5	Best antivirus program	What antivirus program should I download for my computer? Any rec...
10	9	Bitcoin vs. Ethereum	What is the difference between Bitcoin and Ethereum? Which one sho...
	NULL	NULL	NULL

Highlighted row shows the newly inserted tuple into the table.

```
INSERT INTO post_question (qid, uid, resolved, timeposted) VALUES (5, 3, FALSE, '2020-04-22 20:23:59');
```

```
11 • INSERT INTO post_question (qid, uid, resolved, timeposted) VALUES (5, 3, FALSE, '2020-04-22 20:23:59');
```

Result Grid |   Filter Rows: | Export:  | Wrap Cell Content: 

	qid	uid	resolved	timeposted
1	3	0		2013-03-12 01:14:56
2	5	0		2017-02-07 05:31:30
3	4	0		2016-12-12 13:41:17
4	2	0		2011-11-30 18:56:23
6	2	0		2001-06-13 11:44:51
7	9	0		2012-03-14 17:24:23
8	10	0		2007-12-15 03:54:52
9	3	0		2009-08-07 04:59:16
10	5	0		2012-11-01 06:25:00
▶	5	3	0	2020-04-22 20:23:59

(3) Write a query that computes for each user their current status (basic, advanced, or expert status) based on their answers and your own chosen criteria for defining the status.

- Grade is based on the number of likes the answer received, and weight is the weight of those points based on how fast this user posted
 - Therefore, to get the number of points for each user, it is the sum of all their answers' grade*weight.
- Basic users have less than 100 points (automatically set to basic when creating a user), advanced users have more than 100 points, and experts have over 200 points
- These queries first calculates the current points, updates the users' status based on these points, and outputs all the users with their status

```
CREATE VIEW point_table as
```

```
SELECT uid, sum(grade * weight) as points
```

```
From post_answers
```

```
Group by uid;
```

```
UPDATE users, point_table
```

```
SET users.status = 'Advanced'
```

```
Where point_table.points > 100 and users.uid = point_table.uid;
```

```
UPDATE users, point_table
```

```
SET users.status = 'Expert'
```

Where point_table.points > 200 and users.uid = point_table.uid;

SELECT uid, status from users;

uid	status
1	BASIC
2	Expert
3	BASIC
4	BASIC
5	Advanced
7	BASIC
8	BASIC
9	BASIC
10	BASIC

(4) For a given question (say identified by an ID), output all answers to the question in chronological order from first to last. Output the answer text and the time and date when it was posted, and whether an answer was selected as the best answer.

- For this question, the question we are looking for will be qid = 1

Select qid, aid, body, timeposted, best
From answers **join** post_answers **using** (aid)
where qid = 1
Order by timeposted **asc**;

```
19 • Select qid, aid, body, timeposted, best
20   From answers join post_answers using (aid)
21   where qid = 1
22   Order by timeposted asc;
```

Result Grid					Filter Rows:	Export:	Wrap Cell Content:
qid	aid	body	timeposted	best			
1	9	Have you tried riding an airplane?	2021-07-22 01:16:53	0			
1	10	Workout by doing jumping jacks. Maybe you'll jump high enough to fly.	2021-07-22 03:41:18	0			
1	1	You cannot fly. We are bipeds, not birds.	2021-08-20 01:23:51	0			
1	11	If you jump underwater, you're technically you're flying because you push water down like birds push air down.	2023-12-31 16:11:09	1			

(5) For each topic in the topic hierarchy, output the number of questions posted and total number of answers posted within that topic.

- This will include all topics regardless if they have a question or an answer to them

- Subtopics are included with their tid (topic ids rather than stid - their subtopic ids), but with their subtopic name
 - Because the data was unioned together, the snake (subtopic name) is still labeled under tname
- The first 5 rows are the general topics followed by the subtopics
- There are two “Other” subtopics to make sure that even with the same name, the subtopics group based on their subtopic ids (these are considered two different subtopics)

(Select tid, tname, count(distinct qid), count(aid)

From topic left outer join (subtopic join questions using (stid) left outer join post_answers using (qid)) using (tid)

Group by tid, tname)

Union

(Select tid, sname, count(distinct qid), count(aid)

From topic join (subtopic left outer join (questions left outer join post_answers using (qid)) using (stid)) using (tid)

Group by stid, sname);

tid	tname	count(distinct qid)	count(aid)
1	Computer Science	3	2
2	Biology	1	1
3	Mathmatics	2	3
4	Business	1	0
5	Off-topic	3	1
1	Databases	0	0
2	Botany	1	1
3	Algebra 1	2	3
4	Microeconomics	1	0
5	Other	2	1
1	CyberSecurity	2	1
5	Other	1	0
2	Microbiology	0	0
1	Cryptocurrency	1	1

(6) Given a keyword query, output all questions that match the query and that fall into a particular topic, sorted from highest to lowest relevance. (Select and define a suitable form of relevance – you could match the keywords against the query title, the query text, or the query answers, and possibly give different weights to these different fields.)

- The topic in question will be defined as “Computer Science”
- Relevance will be defined as the total number of appearances of the keyword (‘difference’ in this case) discovered in the questions table where the weights are defined as:
 - Title will carry a weight of $0.5 * (\# \text{ of appearances found})$ due to the question title being considered the most important part of a question
 - The body will carry a weight of $0.2 * (\text{number of appearances found in the questions})$

- Related query answers will carry a weight of $0.1 * (\text{number of appearances for each answer})$ due to having the least possibility of relevance for the query
- The views titlecount, bodycount, and answercount count the number of times the keyword appears in the question title, question body, and answers for each question respectively.
 - Keywords are only considered if that exact word is in the text (not a substring of a word)
- Views question_times and answer_times are performing a full outer join on the titlecount, bodycount, and answercount views to ensure all locations where the keyword appears (title, body, answer) are accounted for.
- The results are ordered by the highest weighted count of the keyword.

```
CREATE view titlecount as
  (select qid, (char_length(title) - char_length(REGEXP_REPLACE(title, "\\bdifference\\b",
    'differenc')))) as ttimes
  from questions having ttimes > 0);
```

```
CREATE view bodycount as
  (select qid, (char_length(body) - char_length(REGEXP_REPLACE(body,
    "\\bdifference\\b", 'differenc')))) as btimes
  from questions having btimes > 0);
```

```
CREATE view answercount as
  (SELECT qid, (char_length(body) - char_length(REGEXP_REPLACE(body,
    "\\bdifference\\b", 'differenc')))) as atimes
  from post_answers join answers using (aid) having atimes > 0);
```

```
CREATE view question_times as
  (SELECT qid, ttimes, ifnull(btimes,0) as btimes from titlecount left outer join bodycount
  using (qid))
  UNION
  (SELECT qid, ifnull(ttimes,0) as ttimes, btimes from titlecount right outer join bodycount
  using (qid));
```

```
CREATE view answer_times as
  (SELECT qid, ttimes, btimes, ifnull(atimes,0) as atimes from question_times left outer
  join answercount using (qid))
  UNION
  (SELECT qid, ifnull(ttimes,0) as ttimes, ifnull(btimes,0) as btimes, atimes as atimes from
  question_times right outer join answercount using (qid));
```

```
SELECT qid, title, body, 0.5*ttimes + 0.2*btimes + sum(0.1*atimes) as keyword_count
  from answer_times join questions using (qid) join subtopic using (stid) join topic using
  (tid)
```

Where tname = 'Computer Science'
group by qid, ttimes, btimes
Order by keyword_count desc;

qid	title	body	keyword_count
6	Difference between loops?	What's the difference between a for loop, a while loop, and a do-while loop?	0.7
10	Bitcoin vs. Ethereum	What is these difference between Bitcoin and Ethereum? Which one should I buy into?	0.2

(d) Populate your database with some sample data, and test the queries you have written in part (c). Make sure to input interesting and meaningful data. Limit yourself to a few users, questions, answers, and topics, but make sure there is enough data to generate interesting test cases for the above queries. It is suggested that you design your test data very carefully. Draw and submit a little picture of your tables that fits on one or two pages and that illustrates your test data! (Do not submit a long list of insert statements, but show the resulting tables.) Print out and submit your testing

Topic Table:

	tid	tname	hierarchy
▶	1	Computer Science	1
	2	Biology	1
	3	Mathmatics	1
	4	Business	1
	5	Off-topic	3

Subtopic Table:

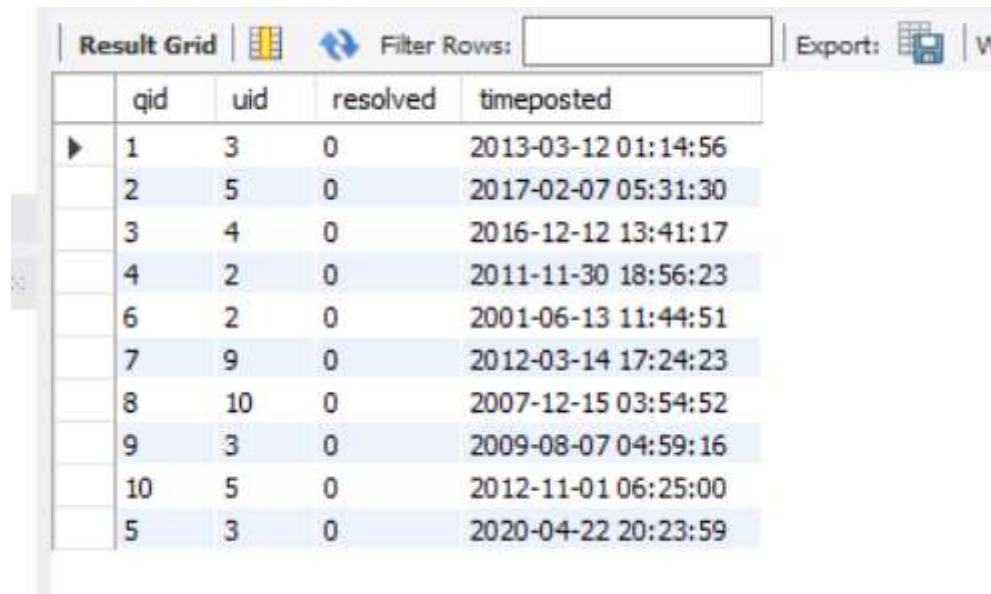
	stid	tid	sname
▶	1	1	Databases
	2	2	Botany
	3	3	Algebra 1
	4	4	Microeconomics
	5	1	CyberSecurity
	6	5	Other
	7	1	Programming
	8	2	Microbiology
	9	1	Cryptocurrency

Questions Table:

	qid	stid	title	body
▶	1	3	HOW TO 1+1	HELP, HOW DO I COMMIT 1+1
	2	4	HOW TO PAY TAXES	Help, the government is inquiring about my car's extended warranty
	3	2	HOW TO GROW PLANT	My crops keep dying when I feed it coca-cola. What am I doing wrong.
	4	5	HOW TO HACK	Hackerman keeps demanding ransom in bittycoin. How do I pay.
	5	6	HOW DO I FLY	I try flapping my arms, but it doesn't work.
	6	7	Difference between ...	What's the difference between a for loop, a while loop, and a do-while loop?
	7	3	Difference between ...	What's the difference between adding 1+1+1 and 1*3?
	8	6	Kingdoms in microbio...	What are the 3 domains of life that live in the microbiology world?
	9	5	Best antivirus program	What antivirus program should I download for my computer? Any recommendations?
	10	9	Bitcoin vs. Ethereum	What is the difference between Bitcoin and Ethereum? Which one should I buy into?

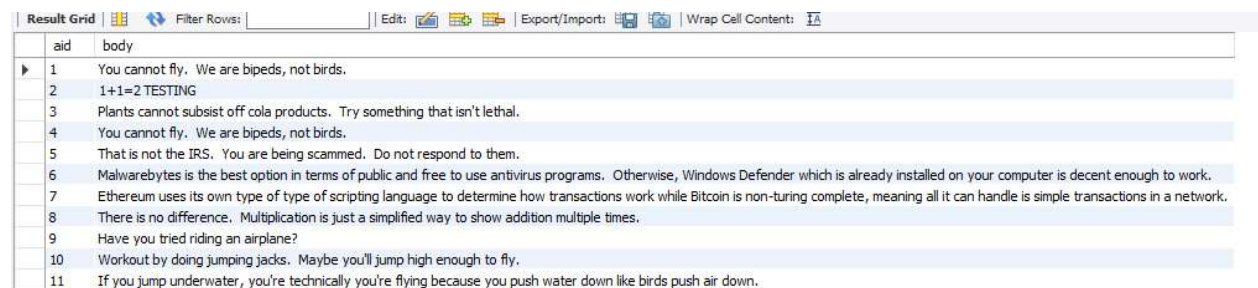
Post_Questions Table:

- Boolean statements are written as 1 and 0 for True and False respectively.



	qid	uid	resolved	timeposted
▶	1	3	0	2013-03-12 01:14:56
	2	5	0	2017-02-07 05:31:30
	3	4	0	2016-12-12 13:41:17
	4	2	0	2011-11-30 18:56:23
	6	2	0	2001-06-13 11:44:51
	7	9	0	2012-03-14 17:24:23
	8	10	0	2007-12-15 03:54:52
	9	3	0	2009-08-07 04:59:16
	10	5	0	2012-11-01 06:25:00
	5	3	0	2020-04-22 20:23:59

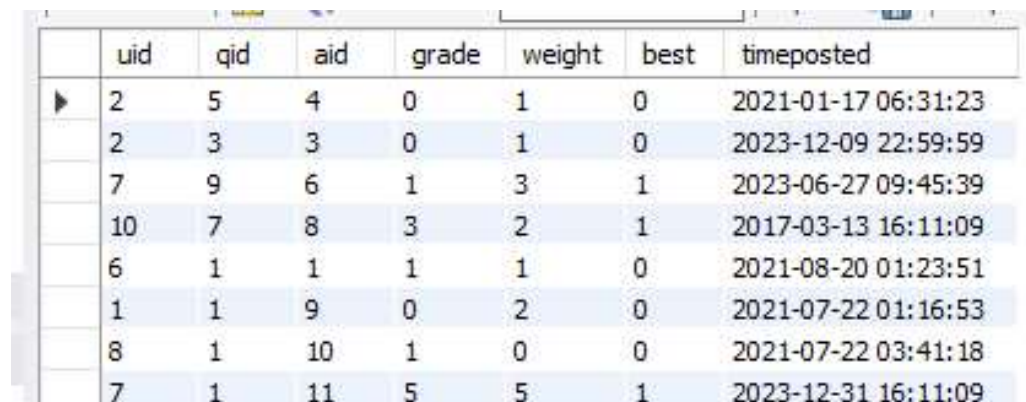
Answers Table:



	aid	body
▶	1	You cannot fly. We are bipeds, not birds.
	2	1+1=2 TESTING
	3	Plants cannot subsist off cola products. Try something that isn't lethal.
	4	You cannot fly. We are bipeds, not birds.
	5	That is not the IRS. You are being scammed. Do not respond to them.
	6	Malwarebytes is the best option in terms of public and free to use antivirus programs. Otherwise, Windows Defender which is already installed on your computer is decent enough to work.
	7	Ethereum uses its own type of type of scripting language to determine how transactions work while Bitcoin is non-turing complete, meaning all it can handle is simple transactions in a network.
	8	There is no difference. Multiplication is just a simplified way to show addition multiple times.
	9	Have you tried riding an airplane?
	10	Workout by doing jumping jacks. Maybe you'll jump high enough to fly.
	11	If you jump underwater, you're technically you're flying because you push water down like birds push air down.

Post_Answers Table:

Booleans are displayed as 1 or 0 in the select query.



	uid	qid	aid	grade	weight	best	timeposted
▶	2	5	4	0	1	0	2021-01-17 06:31:23
	2	3	3	0	1	0	2023-12-09 22:59:59
	7	9	6	1	3	1	2023-06-27 09:45:39
	10	7	8	3	2	1	2017-03-13 16:11:09
	6	1	1	1	1	0	2021-08-20 01:23:51
	1	1	9	0	2	0	2021-07-22 01:16:53
	8	1	10	1	0	0	2021-07-22 03:41:18
	7	1	11	5	5	1	2023-12-31 16:11:09

Likes Table:

Result Grid			
	aid	uid	points
▶	1	3	1
	5	2	1
	8	2	1
	7	1	1
	10	5	1
	9	3	1
	3	4	1
	3	7	1
	3	6	1
	3	3	1

Users Table:

	uid	username	password	profile	status	city	state	country
▶	1	JV4	password	PLACEHOLDER PROFILE	BASIC	Chicago	CO	US
	2	Bobby2	password1	PLACEHOLDER PROFILE	Expert	New York City	NY	US
	3	Mike6	password2	PLACEHOLDER PROFILE	BASIC	Baton Rouge	LA	US
	4	Janet1	password3	PLACEHOLDER PROFILE	BASIC	Houston	TX	US
	5	Daryl9	password4	PLACEHOLDER PROFILE	Advanced	Las Vegas	CA	US
	6	Pippin23	password5	PLACEHOLDER	Basic	LA	CA	US
	7	Darrin4	password6	PLACEHOLDER	BASIC	Austin	TX	US
	8	Mang0	password7	PLACEHOLDER	BASIC	Boston	NY	US
	9	Leffen	password8	PLACEHOLDER	BASIC	Seattle	WA	US
	10	Elphelt	password9	PLACEHOLDER	BASIC	Dallas	TX	US