**MySQL Tutorial**[**https://youtu.be/yPu6qV5byu4**](https://youtu.be/yPu6qV5byu4) *3/12/18 12:07 AM*

**MySQL Beginners Tutorial: MySQL Procedures**[**https://youtu.be/xFNFZM0sS-o**](https://youtu.be/xFNFZM0sS-o) *3/11/18 11:57 PM*

Triggers and Functions:

<https://youtu.be/03ALnWOK5ec>

MySQL Transactions <https://youtu.be/8xpM8pm5yIo>

What a database progress report should have?

What a database progress report should have?

Say your name.

1. What is your project?

2. Show progress!!!!!!!!!!!!!

- Show schema

- Show data

- Auditng/ preprocessing.

- Show social media code.

3. Next steps

How are you thinking on answering the questions in

<https://github.com/nikbearbrown/NEU_COE/blob/master/INFO_6210/Projects/INFO_6210_Database_Project.pdf>

Take Home Exam Part 1 (SQL) - 100 points

All questions are worth 10.0/3  (3.3333) points except #28 which is worth 10 points.

You will be using your project database.

Caveat: Everybody needs to write their own SQL.

Write SQL statements to do the following on  project database:

1. Select columns

2. Filter rows

3. Sort your query

4. Group by an  an attribute

5. Calculate an aggregate function on an attribute of the group.

6. Use DISTINCT keyword will make it so it only returns one instance of each attribute.

7. Create a column that is calculated from other columns.

8. Count all of the null values in a nullable field.

9.  In a text field count all of rows in the columns that contain the letter 'a'.

10. Subselect columns using a subquery.

11. Computationally what is the most expensive operation in the relational data model?

12. Write a function to calculate something relevant to your project.  Show that it works.

13.  When two tables are joined in a relational database what is the resulting data structure?

14. Select and filter some data from a table created by a join.

15.  Why not put all the data in one big table and avoid all of these joins?

16.  Why create views?

17. Select and filter some data from a table created by a view.

18.  Why create temporary tables?

19. Select and filter some data from a table created by a temporary table.

20. Insert some data in to a table.

21. Update some data in a table.

22. Delete some data from a table.

23. Create a  stored routine that does something useful for your project.

24. Create a  stored procedures that does something useful for your project.

25.  Create a index on a non-key attribute. Does it help?

26. Create a trigger.

27. Create a transaction.

28 Explain to an eight year old (i.e. your professor) what are the first  three Normal Forms. (10 points)

SQL vx no SQL fight

[**https://youtu.be/rRoy6I4gKWU**](https://youtu.be/rRoy6I4gKWU)

<https://youtu.be/qI_g07C_Q5I>

MongoDB In 30 Minutes <https://youtu.be/pWbMrx5rVBE>

MongoDB in 5 Minutes with Eliot Horowitz <https://youtu.be/EE8ZTQxa0AM>

MongoDB Tutorial <https://youtu.be/-0X8mr6Q8Ew>

Comparison of PostgreSQL and MongoDB <https://youtu.be/eM7hzKwvTq8>

.timer ON

.mode columns

.headers on

# Clarifications on the Final Project

>>>You need to have data from at least ONE Social Media per member in the group!

>>>You will have 4 Tables other than your data tables which will be scrapped from any Social Media:

**1)Domain Tags:**

What you’ll did for your Assignment 2 already!

**2)Synonyms**

There are various forms of synonyms that you’ll can use:

-Import WordNet library in python, then use wordnet.synsets (Refer [@231](https://piazza.com/class/jc5ig1vl9ud2uy?cid=231))

-Refer to Acronym databases online (https://en.wikipedia.org/wiki/List\_of\_acronyms:\_B)

**3)Mis spelled words**

**Eg: Colour and Color**

For this,

-You can check patterns of words using RegEx

-Levenshtein distance Algorithm (https://en.wikibooks.org/wiki/Algorithm\_Implementation/Strings/Levenshtein\_distance)

**4)Semantic Information**

Almost every real time data that you scrape or retrieve will contain some form of semantic information that you are expected to identify

For e.g.: An Actress can be referred to as a Female Actor. SO here I decide gender to be my semantic information.

An other one could be, A movie to be regarded as a Horror Movie or a Romantic Movie. In these cases I have considered the Genre to be my semantic info.

>>>For the Questions to be answered, Make sure, before you’l write any line of code, you try to put the approach in normal English language sentence and then query accordingly.

i. What are people saying about me (somebody)? **//could be the tags related to that user**

ii. How viral are my posts?

iii. How much influence to my posts have?

iv. What posts are like mine?

v. What users post like me?

vi. Who should I be following? **//1.Similar users and 2. Quality (content of the post- decide on basis of followers that user has so the popularity of that user. If more popular it will have good content!)**

vii. What topics are trending in my domain?

viii. What keywords/ hashtags should I add to my post?

ix. Should I follow somebody back?

x. What is the best time to post?

xi. Should I add and picture or url to my post?

xiii. What’s my reach? **//Distance, the post is seen by only my friends or friends of friends?**

**So you consider those users who are less similar to you that is they follow you but they don’t have similar tags like yours (people who are far, rather than close to you)**

>>>You can write Python scripts to get all this and store in a table. You will answer questions only using a SQL Query!

Hope this helps :)

Please come during TA hours if you have any further questions.

Best,

Urvi