

University of Mary Washington
CPSC350 Applications of Databases Spring 2014

FINAL EXAM

- This exam is due 2:30pm on Tuesday 29 April. You are to submit the exam to the gmail address submit.o.bot. If your exam is delivered after this deadline, it will not be accepted.
- Depending on section, your submission will be in one of three formats:
 - plain text file (required for the sql query section. See instructions in that section)
 - pdf (an option for Part 1)
 - postscript (an option for Part 1)

No other formats will be accepted.

- This test is to be completed **individually** without outside help. This includes no help from peers or the Internet. You are free to use any resources used for the class including the textbook, handouts, and any class notes you made. If there is evidence that any part of the exam was completed with outside help you will receive a 0 for the entire exam.
- To discourage guessing and brain-dump style answers you will receive 20% of the XP for problems you leave completely blank. If you attempt a problem you start at zero XP. By *problem* I mean any numbered problem--subproblems do not count. This means that if you attempt one part of a problem it is best to answer the remaining parts. If you are less than 50XP away from the grade you desire for the class you can submit a blank exam and you will receive 50XP. (**Keep in mind that the minimum score you can receive for the exam is 0**)

Part 1: DB Design (avg50XP; awesome answers: 100XP)

You have the option of selecting one of the following sites

Option 1: Red Blossom Tea

I have a passion for tea. Among my favorite places is redblossomtea.com. An order might look like:

Shipping address: Ron Zacharski Univ. Mary Washington Computer Science Dept. 1301 College Ave. Fredericksburg, VA 22401	Billing address Ron Zacharski 1205 W. Union Ave. Las Cruces, NM 88005
--	--

Shopping Cart:

Name	unit	price	qty	total
Dragonwell – Spring Reserve 2010	2 ounce	\$14.00	1	\$14.00
Wild Leaf, Sheng Pu-erh	4 ounce	\$29.50	2	\$59.00
Qi Zhong Wuyi Oolong	50 grams	\$11.00	1	\$11.00

Look over the website in some detail. Go to the store section. You can ignore the teaware and special items sections of the store and focus on tea.

Option 2: sittercity

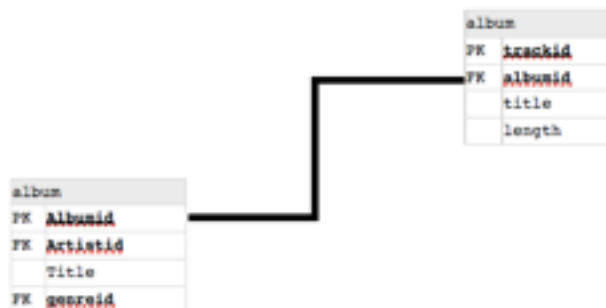
I have used sittercity (<http://www.sittercity.com>) to find dog sitters for my dogs. You can explore much of the site by following the links on the bottom of the home page. Look over the site in detail.

The task

I would like you to design the database to power one of these sites. Tables should be in 3NF.

Warning: To get any reasonable amount of XP for this question I want an intelligent and thoughtful answer.

Imagine that you are at job interview. They give you a few minutes to look over the site, and then ask you to go to the whiteboard and talk through the design. Or better still, imagine that you are going to bid on developing this website and to win the contract you need to pitch your design to someone with my level of expertise. Every 't' does not need to be crossed. **Your answer should include a database schema as described starting on p. 296 of the Head First book.** It should also include **at the bare minimum** one page describing the design. Tables in your schema diagram should follow the following convention:



Where PK = primary key (also indicated by underline and bold) and FK = foreign key (bold).

A complete answer should contain:

1. A one-page relational database diagram as described above
2. A description of the database subject area (what the universe of discourse is).
3. For each table of the design:
 1. A general English description/definition of what the table contains. (not *The user table contains the columns* but *The user table contains general information about the user including login information and location. It does not contain any comments the user made. Those are in table x.*
 2. For each column, the name and definition; any relationships, default values, whether it is a key, whether it is to be indexed.
4. Describe the relationships among tables (i.e., one to many, many-to-many, etc.)
5. initial size of the table and expected growth.

My expectations are high:

Your document should be one you would happily show on a job interview. If I am wowed by your answer, I will give you 100XP. In a hypothetical job situation, if I would keep your design on a shortlist of applications you will get 50XP. Designs of the quality submitted for the Winter Olympics assignment will get at best 25xp. **Please send your document as a pdf or postscript.** No other formats will be accepted..

Part 2: SQL Queries (40 points)

Please generate SQL queries and show the requested results for the following English queries. For this part, please copy your answers to a plain text file labeled part2.sql. For each question you should have a comment identifying the question number and on a separate line the sql query (the query needs to be on a single line). For example:

```
-- 1
select * from table;

-- 2
select username, password, from secretTable where username == 'foo';
```

1. (db: world) Please generate an alphabetical list of languages and the number of countries where that language is spoken. (column names Language and Number of Countries) (3 points)
2. (db: world) What are the worst 5 continents (ordered list please) in terms of life expectancy (show continent name and life expectancy in columns Continent and Life Expectancy). (3 points)
3. (db: world) The average population of an American City (countrycode = 'USA') (column should be labeled 'Average Population') (3 points)
4. (db: world) An alphabetical list of all forms of government (no duplicates please) (3 points)
5. (db: world) The name, country name, continent, and population of the three largest cities in the world ordered by population. (columns City, Country, Continent, Population) (3 points)
6. (db: chinook) An alphabetical list of music artists whose name starts with 'Frank'. column labeled Artist (3 points)

7. (db: chinook) All tracks by the artist Pearl Jam ordered alphabetically by track name. I want columns titled 'artist', 'album' and 'track' (3 points)
8. (db: chinook) A list of artists who have not released any albums, ordered alphabetically. (3 points)
9. (db: movie) What movies was Milla Jovovich in? A single column labeled 'Movie' (3 points)
10. (db: movie) Create a view called costars that contains the names of all their actors and costars. (columns named actors and costars) I don't want actors being costars with themselves and I don't want duplicates. (4 points)
11. (db: movie) Using the view costars from above, how many actors were in films with costars of Kevin Bacon? Column labeled Number of Actors (4 points)
12. (db: movie) Using the view costars from above, I would like an alphabetical list of actors who were both costars of Harrison Ford and Uma Thurman (column called costar) (4 points)
13. (db: movie) A list of movies with the **highest** rating (score). Show movie title and rating. Order alphabetically. (4 points)

Part 3: More SQL Queries: The extremes → easy & hard (50 points)

Points will be deducted for using views.

For this part, please copy your answers to a plain text file labeled part3.sql. For each question you should have a comment identifying the question number and on a separate line the sql query (the query needs to be on a single line).

Note: it is possible to add columns to a result. For example, if my users table looks like:

```
+-----+-----+-----+
| id      | lastname | firstname |
+-----+-----+-----+
| 01      | Smith    | John      |
+-----+-----+-----+
| 02      | Smith    | Mary      |
+-----+-----+-----+
```

And if my query is `select firstname, lastname, 'superuser' from users` I get the table:

```
+-----+-----+-----+
| firstname | lastname | superuser |
+-----+-----+-----+
| John      | Smith    | superuser |
+-----+-----+-----+
| Mary      | Smith    | superuser |
+-----+-----+-----+
```

You will need to use this information for one of the questions below.

1. (db: imdb_small) I want to know how many movies are in the database (column labeled Number of

Movies).

2. (db: imdb_small) I want a list of years with the number of movies released that year, order by the number of movies. The result should look like:

year	Number of Movies
2003	4
1999	4
2000	4
2001	3
1995	2

etc.

3. (db: imdb_small) I want to know the years that had the max movies released ordered by year (**your query should not have a hard coded '3' or '4'**). The result should look like:

year	Number of Movies
1999	4
2000	4
2003	4

4. (db: imdb_small) What is the most popular (most frequent) role?

5. (db: imdb_small) What is the most popular (most frequent) role excluding 'Himself' and the empty string?

6. (db: imdb_small) Who (first name & last) played Trinity in the movie The Matrix (in the db as Matrix, The)?

7. (db: imdb_small) What roles has Kevin Bacon played (order by role)?

8. (db: imdb_small) I would like to know which movies had multiple actors playing the same role. For example, Star Wars had multiple actors playing the role of 'Storm Trooper'. I would like the name of the movie and the name of the role. (columns named Movie and Role)

9. (imdb_small) I would like to see the actors and directors (first names, last names, and roles) for the film Lost in Translation. The role of directors should be listed as 'Director'. Ordered by last name. To be clear, I want one table that looks like:

firstname	lastname	role
-----------	----------	------

Where the role of actors is that given in the table and the role of the director is 'director'

10. I would like a list of movies that have exactly 2 directors. I want the resulting table to look like:

movie	year	first_name	last_name	first_name	last_name
-------	------	------------	-----------	------------	-----------

I want each movie occurring only once in the table. Order by movie name

Part 4: map-reduce 60xp

The problem uses the zipcode json data on our website. Can you write a javascript program that implements Map Reduce that accesses a mongo server to output the smallest city population wise for each state. 0xp for a solution that does not return the correct result. 40xp for a solution that returns the correct result. 70xp for a near optimal solution.