CSC 380/530 — Advanced Database
Take-Home Final Exam (document version 1.0)
SQL, PL/SQL, NoSQL, and MongoDB

- The take-home final exam is due by 11:59:59 PM on Tuesday, December 15, 2015 and must be submitted electronically.
- The take-home final exam will count as 20% of your final course grade.
- The take-home final exam is to be completed **individually**. Do not share your work with anyone else.
- **No late submissions will be accepted.**

Getting Started

For our take-home final exam, there are two parts to complete. First, as with previous assignments, you will write SQL and PL/SQL to answer a series of questions. Second, you will explore the concepts behind NoSQL databases and MongoDB.

Part I: OSPD

For the first part of the final exam, download the `ospd.sql` file from the course website, then execute the script within your Oracle environment. This file contains only one table (called `ospd_words`) that contains 79,339 valid words from the Official Scrabble Player's Dictionary (OSPD). Note that the words in this file are anywhere from two to eight characters long. Also note that words are entirely in lowercase.

Begin by writing SQL queries (no PL/SQL yet) to answer the questions below. As per usual, your SQL code should be as general-purpose as possible (i.e., assume that words may be of any length and that the `ospd_words` table could actually have millions of rows of data).

1. How many words start or end with the letter `'g'` (only show the count)?
2. How many four-letter words are there in the dataset?
3. What words contain both the letter `'w'` and the letter `'x'` (show these words in alphabetical order)?
4. What words contain the letter `'q'` followed by any letter other than `'u'` (show the words in alphabetical order)?

5. Write a single SQL query that shows word-counts for each word-length two through eight; in other words, produce the following output:

```
WORD_LENGTH WORD_COUNT
----------- ----------
          2         94
          3        961
          4       3857
          5       8551
          6      15055
          7      22821
          8      28000
```

6. Write a single SQL query that shows all words that contain the substring `"smart"` (including the word "smart").

7. Write a single SQL query that shows all words that contain the substring `"dumb"` (including the word "dumb").

Next, shift your focus to PL/SQL. Start by creating a PL/SQL package called `word_utils`. Within this package, define and implement PL/SQL functions or procedures that meet the criteria below.

8. Write a PL/SQL function called `is_valid()` that confirms whether input parameter `word` is valid according to the `ospd_words` table (i.e., whether the given `word` is in the dictionary or not). If valid, this function should return a Boolean value of `true`; otherwise, return `false`.

9. Write a PL/SQL function called `num_substring_matches()` that, given input parameter `substring`, counts the number of valid words that contain `substring`. This function must return an integer greater than or equal to zero.

10. Write a PL/SQL procedure called `find_substring_matches()` that, given input parameter `substring`, creates a collection of valid words that contain `substring`. This collection must be an `out` parameter of your procedure and, if non-empty, must be in alphabetical order. Also note that `substring` may contain `'?'` characters that indicate a "wildcard" character that matches any single letter.

    As an example, given substring `"abac"` (case does not matter), the `out` parameter is a collection of the following six words: `abaca`; `abacas`; `abaci`; `aback`; `abacus`; and `abacuses`.

    As another example, given substring `"x?y"` (again, case does not matter), the `out` parameter is a collection of the following seven words: `convexly`; `laxly`; `oxeye`; `oxeyes`; `prolixly`; `reflexly`; and `sixty`.

Part II: NoSQL Databases

Systems that collect and store large amounts of data are growing tremendously. With this explosive growth, the format of such data is increasingly shifting away from decades-old database technologies based on SQL and rigid table structures.

Much of the data collected now is increasingly unstructured or semi-structured data that does not necessarily fit into a rigidly defined database schema.

For the second part of this final exam, your task is to research and study NoSQL databases, including the MongoDB document store. An excellent overview of NoSQL databases is available at the following URL: `http://www.couchbase.com/nosql-resources/what-is-no-sql`.

Your specific task is to review the various NoSQL databases listed at `http://nosql-database.org/`, select MongoDB and **three** other NoSQL databases from this website, then answer the questions below for each of your selections.

Answer each question with three or four sentences (or a bulleted list), perhaps a page for each of your four selections.

If you use other sources, please note their URLs in your answers.

1. How does the given NoSQL database scale to handle extremely large datasets?
2. How does the given NoSQL database handle changes in the datasets that it stores? In other words, how well does the given NoSQL database adapt to change?
3. What are the key differences between the selected NoSQL database and a relational database (i.e., that uses SQL and/or PL/SQL)?
4. What are some of the key limitations of the given NoSQL database?

Submission Instructions

To submit your work, create a single ZIP file (or compressed folder) containing all of your source files and your write-up for Part II. Use your Saint Rose ID (e.g., `goldschmidtd168`) as the name of the ZIP file (i.e., `goldschmidtd168.zip`).

Though entirely optional, you can include a simple `README.txt` file with notes or instructions.

Email your ZIP file to `goldschmidt@gmail.com` (with a subject of "`CSC 380/530 Final Exam`").