CSC 455: Database Processing for Large-Scale Analytics
Assignment 2

Due Thursday, April 20th

Supplemental reading: SQL reference book *Oracle 12c SQL* by Price, **ISBN 9780071799355** (available in Books 24x7 DePaul online library as eBook). Sections:
Pp.31-34, "Performing Arithmetic"
pp.35-40, "Combining Column Output Using Concatentation", "Null Values", "Displaying Distinct Rows", "Comparing Values", "Using the LIKE Operator"
pp. 43- 46, "Using the Logical Operators", "Sorting Rows Using the ORDER BY Clause"

If you want a normalization reference, you can read Section 14.6 in "Information modeling and relational databases", **ISBN 9780123735683** in the online library (although we only need to consider 1NF/2NF/3NF).

Part 1

You are given a following schema in 1NF:
(First, Last, Address, Job, Salary, Assistant) and the following set of functional dependencies:

First, Last → Address
Job → Salary, Assistant

Decompose the schema to make sure it is in Third Normal Form (3NF).

Write SQL DDL and SQL INSERT statements to create the 3NF tables and load the data provided in the text file data_hw2.txt (posted together with this assignment). Make sure that NULL in the data file is loaded as database NULL and not as 'NULL' string. You can use your python code from Assignment 1 to generate these INSERT statements.

Part 2

Write a python script that is going to create your tables from Part 1 and populate them with data automatically. Use sqlite3 database as shown in class and make the necessary data type changes in your DDL from part1 (e.g., NUMBER(5,0)→INTEGER, NUMBER(5,2)→REAL).
Your python code must successfully load data into all tables (however many tables you have created in Part-1). Note that you are loading data into SQLite, not into Oracle (the process is the same, but installing cx_Oracle for python is somewhat complicated). Your python code must read the input file data_hw2.txt directly and automatically load the data. Manually creating INSERT SQL statements will not count for credit in this part.

Part 3

You were hired to do some data analysis for a local zoo. Below is the data table, including the necessary constraints and all the insert statements to populate the database.

```
-- Drop all the tables to clean up
DROP TABLE Animal;

-- ACategory: Animal category 'common', 'rare', 'exotic'.  May be NULL
-- TimeToFeed: Time it takes to feed the animal (hours)
CREATE TABLE Animal
(
  AID       NUMBER(3, 0),
  AName     VARCHAR2(30) NOT NULL,
  ACategory VARCHAR2(18),

  TimeToFeed NUMBER(4,2),

  CONSTRAINT Animal_PK
    PRIMARY KEY(AID)
);
INSERT INTO Animal VALUES(1, 'Galapagos Penguin', 'exotic', 0.5);
INSERT INTO Animal VALUES(2, 'Emperor Penguin', 'rare', 0.75);
INSERT INTO Animal VALUES(3, 'Sri Lankan sloth bear', 'exotic', 2.5);
INSERT INTO Animal VALUES(4, 'Grizzly bear', 'common', 3.0);
INSERT INTO Animal VALUES(5, 'Giant Panda bear', 'exotic', 1.5);
INSERT INTO Animal VALUES(6, 'Florida black bear', 'rare', 1.75);
INSERT INTO Animal VALUES(7, 'Siberian tiger', 'rare', 3.5);
INSERT INTO Animal VALUES(8, 'Bengal tiger', 'common', 2.75);
INSERT INTO Animal VALUES(9, 'South China tiger', 'exotic', 2.25);
INSERT INTO Animal VALUES(10, 'Alpaca', 'common', 0.25);
INSERT INTO Animal VALUES(11, 'Llama', NULL, 3.5);
```

Since none of the managers in the zoo know SQL, it is up to you to write the queries to answer the following list of questions.

1. Find all the animals (their names) that take less than 1.5 hours to feed.
2. Find all the rare animals and sort the query output by feeding time (from small to large)
3. Find the animal names and categories for animals related to a bear (hint: use the LIKE operator)
4. Return the listings for all animals whose rarity is missing in the database
5. Find the rarity rating of all animals that require between 1 and 2.5 hours to be fed
6. Find the names of the animals that are related to the tiger and are not common
7. Find the names of the animals that are <u>not</u> related to the tiger
8. Find the minimum and maximum feeding time amongst all the animals in the zoo
9. Find the average feeding time for all of the rare animals

<u>Be sure that your name and “Assignment 2” appear at the top of your submitted file</u>.