Ass1: Query Implementation using RDBM System and Elastic Search
(Issue: Jan 09, Due: Jan 25)

- TA: Trishla Shah trishla@dal.ca
- Ass1-Tut1: Jan 15, 4:05-5:25PM; Ass1-Tut2: Jan 22, 4:05-5:25PM, Chemistry 226
- Ass1-help Hours: Jan 11, 18, 24, 2:00-3:30PM, CS 233

1. **Objectives:**
    1) To review conventional Relational Database Systems
    2) To learn Infrastructure Services on a Cloud System
    3) To learn Distributed Database concepts and implementation
    4) To learn installation of Relational Database on a Cloud VM

2. **Tasks:**
    1) Open a personal account on IBM Bluemix cloud service. (* If you prefer to use any other option, a justification must be provided.)
    2) Create a DB instance running IBM's DBMS (DB2) on Bluemix cloud service.
        a. You are given a dataset containing data instances of "Bus Routes", "Bus Stops" and "Bus Stop Times".
        b. Create three tables : Trips, Stops and Stoptimes and Load data into the tables from the given data .csv files.
    3) Perform following queries:
        a. Find all buses for a particular Bus Stop
           1. Input: Bus Stop Name
           2. Output: List of all buses, response time for the search query
        b. Find buses between two time ranges
           1. Input: Time Range 1 (hh:mm:ss), Time Range 2 (hh:mm:ss)
           2. Output: List of all buses, response time for the search query
        c. List all the bus-stops in sequence of a particular bus on a particular route (Herein, route information signifies all the stops and their sequence made by the bus running on the provided route). (Hint: For particular route use route_id and for particular bus use trip_headsign)
           1. Input: Bus Name, Route Name
           2. Output: List of all routes, response time for the search query
        d. Find top 3 bus stops that are the busiest throughout the day. (Hint: The bus stops with high volume of bus routes).
           1. Input: None
           2. Output: List of Bus Name, response time for the search query
       Note: - During loading set maximum no. of warnings to 999999999 instead of 1000. This will allow to load the database upto the maximum warnings.
           - Some rows in the database might get rejected due to noise or other errors present. That must be ignored.

4) Use IBM Bluemix's Compose for Elasticsearch database. Import data for Bus Route. Run search query on all three documents.

   Note: It is not necessary to use nested or join queries for elasticsearch. You can write separate sub queries and use hardcoded input (Hint: Output of first sub query will be become input for second sub query and so on.)

5) Observe the difference in performance and search quality. Explain why one would use Elastic search over the local search or vice versa.

6) Write a report including the following sections:
   a. Task Description: Present the business scenario (i.e., the application and the requirements), and the DB (provide an instance of the database, and the reference source, etc.).
   b. Relational Database Design: Provide an overview of schema for the database. Please mention Primary Key, Foreign Keys and Indexes if any. You can present this in simple text or an ERD Diagram.
   c. Application Queries & Test Results: Please provide details on experimentation performed on relational database and elastic search. Please mention your search queries and response time.
   d. Summary: Provide a summary of your work & observations on the application (i.e. the original DB) and the experience of using the software tools (i.e. your comments & recommendations, etc.).

3. Submit your Ass1 report electronically:
1) Please use Bright Space to submit your assignment
2) In addition to the report please also submit queries.
3) Please also provide any scripts that you used for implementation of Relational Database on Cloud and/or local system (if any).
4) Submit all items within one Zip file

* Plagiarism and Intellectual Honesty: (http://plagiarism.dal.ca)
Dalhousie University defines "plagiarism as the presentation of the work of another author in such a way as to give one's reader reason to think it to be one's own."
Plagiarism is considered a serious academic offense which may lead to loss of credit, suspension or expulsion from the University, or even the revocation of a degree.