

Assignment 1: Query Implementation using Relational Database and Elastic Search

(Issue: Jan 10, Due: Jan 24, 11:59PM)

- TA: Abhinav Kalra, Abhinav.Kalra@Dal.Ca
- Tutorial: Jan 11, 1:00-2:30 PM, Room: 134
- Help Hours: Wed 1:00-2:30PM, Room 134; Fri, 1:00-2:30PM, Room: 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) Create a personal account on a cloud hosting that offers Infrastructure as a service. You are recommended to use Amazon Web Services as Cloud Service Provider. (*Using any other option, such as Microsoft Azure, Digital Ocean, etc., a justification must be provided.)
- 2) Create a virtual machine on your choice of selected Cloud Service.
- 3) Choose MySQL as your RDBMS and install it on your local machine. (*Using any other RDBMS, a justification must be provided.)
- 4) You are provided with a dataset containing Bus Routes, Bus Stops and Bus Stop Times. However, you can also download similar information from the source available below:
 - a. <http://catalogue.hrm.opendata.arcgis.com/>
- 5) Import the data set into available relational database you installed in step 3.
- 6) Use a simple UI application that will perform below searches on the data set. UI application should be able to take text as input and should either display or store results file on disk.
 - 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. Find route information of a particular bus on a particular route
 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 in terms of bus routes. (Hint: The bus stops with high volume of bus routes and close time gaps would be considered as busiest).
 1. Input: None
 2. Output: List of Bus Name, response time for the search query
- 7) Download Elastic Search database from the source available below and import data for Bus Route (<https://www.elastic.co/products/elasticsearch>). Run search query on all three documents. You are free to use Cloud hosting for Elastic search.
- 8) Repeat Step 6 (search queries) using Elastic Search that you created in Step 7 on the cloud. Observe the difference in performance and search quality. Explain why one would use Elastic search over the local search or vice versa.
- 9) **BONUS:** Scale Elastic Search to make use of 2 nodes and multiple Shards and observe performance difference. Observe scalability and comment on performance.
- 10) 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: Provide description of the search text query and provide response time in (milliseconds) that it takes to execute a single query using a single node on web server and database server. Please comment on performance between local system and cloud system.
 - d. Test Results: Please provide details on experimentation performed on relational database and elastic search. Please mention your search queries and response time. It would be useful to have charts as a form of representation.
 - e. Summary: Provide a summary of your work & observations on the application (i.e. the original DB), the developed user interface 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 code for elastic search implementation and queries.
- 3) Please also provide any scripts that you used for implementation of Relational Database on Cloud and/or local system

*** 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.