Assignment #3

CSCI 5408 (Data Management, Warehousing, Analytics)
Faculty of Computer Science, Dalhousie University

Date Given: Feb 14, 2022
Due Date: Feb 27, 2022 at 11:59 pm

Late Submissions are not accepted and will result in a late penalty of 10% deductions / day in the assignment.

Objective:

- The objective of this assignment is to perform Descriptive, and exploratory research and analysis

Plagiarism Policy:

- This assignment is an individual task. Collaboration of any type amounts to a violation of the academic integrity policy and will be reported to the AIO.
- Content cannot be copied verbatim from any source(s). Please understand the concept and write in your own words. In addition, cite the actual source. Failing to do so will be considered as plagiarism and/or cheating.
- The Dalhousie Academic Integrity policy applies to all material submitted as part of this course. Please understand the policy, which is available at: https://www.dal.ca/dept/university_secretariat/academic-integrity.html

Assignment Rubric

	Excellent (25%)	Proficient (15%)	Marginal (5%)	Unacceptable (0%)
Completeness including Citation	All required tasks are completed	Submission highlights tasks completion. However, missed some tasks in between, which created a disconnection	Some tasks are completed, which are disjoint in nature.	Incorrect and irrelevant
Correctness	All parts of the given tasks are correct	Most of the given tasks are correct However, some portions need minor modifications	Most of the given tasks are incorrect. The submission requires major modifications.	Incorrect and unacceptable
Novelty	The submission contains novel contribution in key segments, which is a clear indication	The submission lacks novel contributions. There are some evidences of novelty, however, it is not significant	The submission does not contain novel contributions. However, there is	There is no novelty

	of application knowledge		an evidence of some effort	
Clarity	The written or graphical materials, and developed applications provide a clear picture of the concept, and highlights the clarity	The written or graphical materials and developed applications do not show clear picture of the concept. There is room for improvement	The written or graphical materials, and developed applications fail to prove the clarity. Background knowledge is needed	Failed to prove the clarity. Need proper background knowledge to perform the tasks

Citation:
McKinney, B. (2018). The impact of program-wide discussion board grading rubrics on students' and faculty satisfaction. Online Learning, 22(2), 289-299.

Tasks

Problem #1

1. Read the given paper and write summary (Maximum 2 pages in your own words - Do not copy any content verbatim).
2. In addition, explore** and report, if the paper has any scope of improvements in terms of technical details or concepts.
** Due to time constraint, and limitation of formal research background, you do not need to explore in details the scientific equations or parameters presented in the paper.
   Visit and Login to https://dal.ca.libguides.com/ieee and search the following papers.
3. Do not forget to provide citation in proper format for the given paper and for any other supporting materials that you may use as reference.

Database Administration and Security Reading

G. Vonitsanos, E. Dritsas, A. Kanavos, P. Mylonas and S. Sioutas, "Security and Privacy Solutions associated with NoSQL Data Stores," *2020 15th International Workshop on Semantic and Social Media Adaptation and Personalization (SMA*, 2020, pp. 1-5, doi: 10.1109/SMAP49528.2020.9248442.
Paper Link: https://ieeexplore-ieee-org.ezproxy.library.dal.ca/document/9248442

Problem #1 Submission Format:

Submit one single PDF file

Problem #2 (MySQL and Java – Implementation of 2 phase locking)

1. Visit the website and download the dataset https://www.kaggle.com/johnharshith/hollywood-theatrical-market-synopsis-1995-to-2021
2. Create a database in your local MySQL server, and import the CSV files to your tables. [Each CSV file should represent a table in your database].
   I.   You can perform cleaning and formatting of the CSV same as Assignment 1, or Assignment 2. (using few sentences and screenshots, exhibit what kind of cleaning or formatting you have performed)
   II. If CSV files contain huge volume of data, you can consider max 3000 data points

3. Write a Java program (without using any packages/libraries etc.) to implement concurrent transaction with 2-phase locking protocol – This program will perform 3 major operations using 3 methods.

   You can initiate two Java priority threads representing two concurrent transactions

Table 2.1: Highlights the instructions and interleaving sequence

T1	T2
Select 'TICKET SOLD' from 'AnnualTicketSales' where 'YEAR'=2011;	
	Update 'AnnualTicketSales' set 'TICKET SOLD' = 0 where 'YEAR'=2011; Select 'TICKET SOLD' from 'AnnualTicketSales' where 'YEAR'=2011;
Update 'AnnualTicketSales' set 'TICKET SOLD' = 0 where 'YEAR'=2011; Commit;	
	Commit;

Method 1: acquire_locks() //In this method, each transaction will get the lock requested. Depending on arrival the transaction will get the lock

Method 2: transaction() //This method will be called for each transaction that acquired the lock; In this method a transaction will perform the execution of SQL statement and will select or perform update in the MySQL. Make sure to use autocommit =false in your code; This will be set to true only if commit is reached.

Method 3: release_ locks()//In this method, each transaction will release the lock acquired before

---

Problem #2 Submission Format:

Submit one single PDF file; SQL Dump (Structure + Value); Java Code; screenshots

---

1. Read the given paper and write a summary within 1 page. (**Do not copy any content verbatim**).
2. Considering the concept discussed in the paper, you need to draw **a high-level architecture (block diagram)** to present a data management system for **Canada's winter storm data** (I am not providing any dataset for winter storm. This is a scenario you need to assume; You can use weather site, news, articles, blogs etc. to gain some insight). To draw the architecture, you can use draw.io or power point or similar drawing tools.
3. **Write a ½ page description of your block diagram/ architecture.**
4. Do not forget to provide citation in proper format for the given paper and for any other supporting materials that you may use as reference.

Data Management of Massive Data Reading

H. Wang, X. Tang, S. Shi and F. Ye, "Research on the Construction of Data Management System of Massive Satellite Images," *2018 International Workshop on Big Geospatial Data and Data Science (BGDDS)*, 2018, pp. 1-4, doi: 10.1109/BGDDS.2018.8626845.
Paper Link: https://ieeexplore-ieee-org.ezproxy.library.dal.ca/document/8626845

Problem #3 Submission Format:

Submit one single PDF file – must contain summary, block diagram/architecture, description