Assignment 6 & 7 – Query Optimization

Dates of work: Sep 20, 2022, Sep 27, 2022;

Submission of complete assignment (code + report of analyses): Oct 18, 2022;

Viva: Oct 22, 2022 & Oct 23, 2022; Maximum Points: 330

Questions:

1. Consider the following partial tables of an ordered library catalogue:

BOOK

Author_ID	Book_ID	Author_Name	Book
An_Ch_0103	Aest_AC_0103	Anjan Chatterjee	The Aesthetic Brain
An_Da_0104	Self_AD_0104	Antonio Damasio	Self Comes to Mind
Ca_Sa_0319	Beyo_CS_0319	Carl Safina	Beyond Words: What Animals Think and Feel
Ca_Sa_0319	Song_CS_0319	Carl Safina	Song for the Blue Ocean
Jo_Ro_1018	Deat_JR_1018	Joanne K. Rowling	Deathly Hallows_Harry Potter
Jo_Ro_1018	Fant_JR_1018	Joanne K. Rowling	Fantastic Beasts and Where to Find Them
Jo_Ro_1018	Gobl_JR_1018	Joanne K. Rowling	Goblet of Fire_Harry Potter
Jo_Ro_1018	Phil_JR_1018	Joanne K. Rowling	Philosopher's Stone_Harry Potter
Jo_Ro_1018	Pris_JR_1018	Joanne K. Rowling	Prisoner of Azkaban_Harry Potter
La_Ch_1203	Mind_LC_1203	Lars Chittka	The Mind of a Bee
Ma_Mi_1313	Emot_MM_1313	Marvin Minsky	Emotion Machine
Ma_Mi_1313	Soci_MM_1313	Marvin Minsky	Society of Mind
Pe_Wo_1623	Aunt_PW_1623	Pelham G. Wodehouse	Aunts Aren't Gentlemen
Pe_Wo_1623	Wode_PW_1623	Pelham G. Wodehouse	Wodehouse at the Wicket
Vi_Ra_2218	Emer_VR_2218	Vilayanur Ramachandran	The Emerging Mind
Vi_Ra_2218	Phan_VR_2218	Vilayanur Ramachandran	Phantoms in the Brain

AUTHOR

Author_ID	Author_Name
An_Ch_0103	Anjan Chatterjee
An_Da_0104	Antonio Damasio
Ca_Sa_0319	Carl Safina
Jo_Ro_1018	Joanne K. Rowling
La_Ch_1203	Lars Chittka
Ma_Mi_1313	Marvin Minsky
Pe_Wo_1623	Pelham G. Wodehouse
Vi_Ra_2218	Vilayanur Ramachandran

BOOK PURCHASE

Book_ID	Book	Purchase_Date	Quantity
Aest_AC_0103	The Aesthetic Brain	Sep 5, 2022	1
Self_AD_0104	Self Comes to Mind	Sep 5, 2022	1
Beyo_CS_0319	Beyond Words: What Animals Think and Feel	Sep 5, 2022	2
Song_CS_0319	Song for the Blue Ocean	Sep 6, 2022	2
Deat_JR_1018	Deathly Hallows_Harry Potter	Sep 7, 2022	5
Fant_JR_1018	Fantastic Beasts and Where to Find Them	Sep 6, 2022	5
Gobl_JR_1018	Goblet of Fire_Harry Potter	Sep 5, 2022	5
Phil_JR_1018	Philosopher's Stone_Harry Potter	Sep 5, 2022	5
Pris_JR_1018	Prisoner of Azkaban_Harry Potter	Sep 5, 2022	5
Mind_LC_1203	The Mind of a Bee	Sep 6, 2022	2
Emot_MM_1313	Emotion Machine	Sep 5, 2022	1
Soci_MM_1313	Society of Mind	Sep 6, 2022	1
Aunt_PW_1623	Aunts Aren't Gentlemen	Sep 7, 2022	4
Wode_PW_1623	Wodehouse at the Wicket	Sep 5, 2022	4

Emer_VR_2218	The Emerging Mind	Sep 5, 2022	1
Phan_VR_2218	Phantoms in the Brain	Sep 6, 2022	3

Note: Majority of searches in the catalogue involve Author name and/or Book name

[All codes must be written in C / C++]

Q1.	a. Is the given database normalized? Would you want to normalize the given database any further? (Points: 10 + 10) b. Identify the different indexes that you would use for the tables in the database. Justify your answer. (Points: 10 + 10) Use the normalized database and the indexes to answer the following questions	[40]
Q2.	Use a 5-bucket extendible hash + doubly linked lists (as the core database arrangement structure for all tables) to compare between the different variations of optimization of the following queries on the given tables [Answers must begin with the Relational Algebra expression, followed by the different Query-trees and execution of queries as per the query plans]: a. Retrieve names of books written by 'Carl Safina' b. Retrieve book names and author details of all books written by authors with names beginning with 'A' or 'P' c. Retrieve all books with >= 5 copies d. Retrieve author_names whose books have been purchased across all dates available on the purchase table e. For the books that have just a single copy in the database, retrieve the author_names (Points: 10 for choice of hash function on index + (10 * 5) for RA expressions and query tree designs + (10 * 5) for comparisons in designs per query)	[110]
Q3.	Use a 5-bucket linear hash + doubly linked lists (as the core database arrangement structure for all tables) to compare between the different variations of optimization of the following queries on the given tables [Answers must begin with the Relational Algebra expression, followed by the different Query-trees and execution of queries as per the query plans]: a. Retrieve names of books written by 'Carl Safina' b. Retrieve book names and author details of all books written by authors with names beginning with 'A' or 'P' c. Retrieve all books with >= 5 copies d. Retrieve author_names whose books have been purchased across all dates available on the purchase table	[110]

	e. For the books that have just a single copy in the database, retrieve the author_names (Points: 10 for choice of hash function on index + (10 * 5) for RA expressions and query tree designs + (10 * 5) for comparisons in designs per query)	
Q4.	Compare between execution times for all queries in Q.2 and Q.3 [Comment on the role of the data-structures used, and complexity of the different query plans.]	[50]
** Bonus points: Use of B+-tree data structures in sync with the hashing mechanisms for Q.2 and Q.3 (10 + 10) **		[20]

Assessment Rubric for Submitted Work - Evaluation per answer (Viva & Analyses):

Proper understanding of question and work done accordingly:	80 - 95%
Attempt to work beyond what has been asked, with in-depth understanding: (Definite contender for full score for question)	95 - 100%
Vague understanding, bursts of in-depth answers:	70 - 80%
Vague understanding, bursts of broad conceptual answers:	50 - 70%
Weird hash of work submitted, some understanding:	40 - 50%
No understanding, just work submitted: (Probable plagiarism)	0

Submission Rubric:

Within 2 days of Deadline:	No penalty
Within 5 days of Deadline:	30% penalty
Within 7 days of Deadline:	50% penalty
After 7 days of Deadline:	Will not be evaluated