

# CSCI 620 Introduction of Big Data

## Homework 2 SQL due Feb 19<sup>th</sup>

Guo, Zizhun

[zg2808@cs.rit.edu](mailto:zg2808@cs.rit.edu)

This homework is built based on **MySQL Workbench**, please connect to server, and import **.sql** files (*HW2Q1.sql*, *HW2Q2.sql*) to test it. For a fresh start, tap “ctrl+shift+enter” in file, and “ctrl + enter” for each query.

Question 1:

	book	year	publisher	price	num
▶	1000	1999	Rowan Boyle	14	200
	1001	2010	Paul Gray	32	320
	1002	1994	Holly Wilson	55	50
	1003	1995	Daphne Merkin	20	250
	1004	1996	Ewan Blake	78	80
	1005	1999	Yan Lee	30	300
	1006	2003	Sophie Wong	99	20
	1007	1992	Rowan Boyle	12	400
	1008	1999	Meimei Han	32	150
	1009	2003	Daphne Merkin	80	30
•	NULL	NULL	NULL	NULL	NULL

BookPublish

BookReview

	book	reviewer	score
▶	1000	Cleo Carson	4
	1000	Yu Qiao	4
	1001	Joel Reed	5
	1002	Yang Song	5
	1003	Daphne Merkin	4
	1003	Paul Gray	4
	1004	Cleo Carson	4
	1004	Meimei Han	5
	1004	Paula Bradford	3
	1005	Paul Gray	5
	1006	Iqra Whitehead	4
	1007	Cleo Carson	4
	1008	Holly Wilson	3
	1009	Daphne Merkin	5
•	NULL	NULL	NULL

	book	author	earnings
▶	1000	Felix mendez	20000
	1000	Holly Wilson	35000
	1001	Cleo Carson	50000
	1002	Joel Reed	40000
	1003	Qiao Lin	80000
	1004	Ewan Blake	15000
	1004	Rowan Boyle	30000
	1004	Stephanie Crawford	40000
	1005	Meimei Han	30000
	1006	Kira Fowler	35000
	1007	Iqra Whitehead	15000
	1008	Yang Song	25000
	1009	Kira Fowler	20000
•	NULL	NULL	NULL

BookAuthor

BookReference

	book	referencedBook
▶	1009	1000
	1003	1001
	1006	1001
	1000	1002
	1002	1002
	1008	1002
	1005	1003
	1004	1006
	1001	1009
	1007	1009
•	NULL	NULL

1) Find all the books published in 1999 and reviewed by Paul.

Result Grid	Filter Rows:	Export:	Wrap Cell Content:
book			
1005			

2) Find all the books reviewed by both Paul Gray and Daphne Merkin.

Result Grid	Filter Rows:	Export:	Wrap Cell Content:
book			
1003			

3) List each author's total earnings in 1995-2000 in descending order.

Result Grid	Filter Rows:	Export:	Wrap Cell Content:
author			
Qiao Lin	80000		
Stephanie Crawford	40000		
Holly Wilson	35000		
Meimei Han	30000		
Rowan Boyle	30000		
Yang Song	25000		
Felix mendez	20000		
Ewan Blake	15000		

Question 2:

Course

	number	title	credits	syllabus
▶	101	A1	3	610
	102	A2	3	620
	103	A3	3	630
	104	A4	3	640
	105	A5	3	650
	201	B1	3	610
	202	B2	3	620
	203	B3	3	630
	204	B4	3	640
	205	B5	3	650
•	NULL	NULL	NULL	NULL

CourseOffering

	number	year	semester	section_number	classroom
▶	101	2018	fall	1	3010
	102	2018	fall	2	4002
	103	2019	spring	1	2006
	104	2018	fall	2	4020
	105	2019	spring	1	1030
	201	2018	fall	2	2030
	202	2019	spring	1	3030
	203	2018	fall	2	1010
	204	2019	spring	1	1010
	205	2018	fall	1	3010
•	NULL	NULL	NULL	NULL	NULL

Student

	student_id	name	department
▶	7527001	David Lee	Computer Science
	7527002	Cleo Carson	Computer Science
	7527003	Yu Qiao	Computer Science
	7527004	Joel Reed	Computer Science
	7527005	Paula Bradford	Computer Science
	7527006	Paul Gray	Electrical Engineering
	7527007	Iqra Whitehead	Electrical Engineering
	7527008	Holly Wilson	Electrical Engineering
	7527009	Daphne Merkin	Electrical Engineering
	7527010	Sophie Wong	Electrical Engineering
•	NULL	NULL	NULL

CourseOffering\_timing

	number	year	semester	section_number	timing
▶	101	2018	fall	1	900
	102	2018	fall	2	800
	103	2019	spring	1	1200
	104	2018	fall	2	1400
	105	2019	spring	1	1000
	201	2018	fall	2	1500
	202	2019	spring	1	900
	203	2018	fall	2	800
	204	2019	spring	1	1000
	205	2018	fall	1	1400
•	NULL	NULL	NULL	NULL	NULL

Teaches

	instructor_id	name	department	title
▶	12301	Ellis Myers	Computer Science	Professor
	12302	Bethany Holmes	Computer Science	Professor
	12303	Leal Cox	Computer Science	Associate Professor
	12304	Wolfe Torres	Computer Science	Associate, Professor
	12305	Kalie Ross	Computer Science	Assitant Professor
	12306	Jonh Hill	Electrical Engineering	Professor
	12307	Redford Bell	Electrical Engineering	Professor
	12308	Aleksa Thomas	Electrical Engineering	Associate Professor
	12309	Leland Rogers	Electrical Engineering	Associate Professor
	12310	Ladawn Davis	Electrical Engineering	Assitant Professor
•	NULL	NULL	NULL	NULL

Prerequisite

	course_number	prerequisite_number
▶	101	102
	102	103
	101	104
	103	104
	104	105
	201	202
	202	203
	201	204
	203	204
	204	205
•	NULL	NULL

Instructor

	instructor_id	number	year	semester	section_number
▶	12301	101	2018	fall	1
	12302	102	2018	fall	2
	12303	103	2019	spring	1
	12304	104	2018	fall	2
	12305	105	2019	spring	1
	12306	201	2018	fall	2
	12307	202	2019	spring	1
	12308	203	2018	fall	2
	12309	204	2019	spring	1
	12310	205	2018	fall	1
*	NULL	NULL	NULL	NULL	NULL

Enrolls

	student_id	number	year	semester	section_number	grade
▶	7527001	101	2018	fall	1	80
	7527001	102	2018	fall	2	78
	7527002	101	2018	fall	1	92
	7527002	102	2018	fall	2	73
	7527003	103	2019	spring	1	87
	7527003	104	2018	fall	2	61
	7527004	103	2019	spring	1	89
	7527004	104	2018	fall	2	78
	7527005	103	2019	spring	1	83
	7527005	105	2019	spring	1	79
	7527006	201	2018	fall	2	95
	7527006	202	2019	spring	1	76
	7527007	201	2018	fall	2	94
	7527007	202	2019	spring	1	68
	7527008	203	2018	fall	2	83
	7527008	204	2019	spring	1	79
	7527009	203	2018	fall	2	82
	7527009	204	2019	spring	1	92
	7527010	203	2018	fall	2	90
	7527010	205	2018	fall	1	83
*	NULL	NULL	NULL	NULL	NULL	NULL

1) Retrieve student\_ids, names of students with a major in "Computer Science".

Result Grid	Filter Rows:	Edit:	Export/Import:	Wrap Cell Content:
student_id	name			
7527001	David Lee			
7527002	Cleo Carson			
7527003	Yu Qiao			
7527004	Joel Reed			
7527005	Paula Bradford			
HULL	HULL			

2) Retrieve student\_ids of students who have taken a class with an instructor named "John".

Result Grid	Filter Rows:	Export:	Wrap Cell Content:
student_id			
7527006			
7527007			

3) Retrieve student\_ids of students who have never taken a class with an instructor named "John".

Result Grid	Filter Rows:	Edit:	Export/Import:	Wrap Cell Content:
student_id				
7527001				
7527002				
7527003				
7527004				
7527005				
7527008				
7527009				
7527010				
HULL				

4) Retrieve department names and the number of students with that department major.

Result Grid	Filter Rows:	Export:	Wrap Cell Content:
count(s.student_id)	department		
5	Computer Science		
5	Electrical Engineering		

5) Retrieve names of instructors teaching Computer Science courses, the sections (course number, section number, year, semester) they are teaching, and the total number of students in the sections.

Result Grid

Filter Rows:

Export:

Wrap Cell Content:

	Instructor's name	number	year	semester	section_number	Number of students
▶	Ellis Myers	101	2018	fall	1	2
	Bethany Holmes	102	2018	fall	2	2
	Leal Cox	103	2019	spring	1	3
	Wolfe Torres	104	2018	fall	2	2
	Kalie Ross	105	2019	spring	1	1