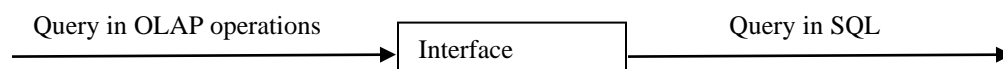


Assignment in 6 Parts for Data Warehouse Course  
Fall 2023

Purpose: Design and manipulation of a data warehouse.

Problem: The University of Savannah at Richmond Hill (USRH) has extracted data from three data sets about her graduates for several years (Data are attached at the end of the assignment). The authorities of the university want you to design a *graduates* data warehouse. This data warehouse facilitates the study of the number of graduates for different majors, degrees, colleges, GPAs, semesters, and/or any combination of them. To satisfy the needs of the authorities do:

1. Integrate the extracted data as one source of data. Take care of the format and semantic conflicts. Some information about the three data sources obtained from the data dictionaries. If you have a question, then you can interview me as the representative of the USRH via zoom. Explanations of all the actions that you have taken and their justifications must be included in a word document. (Name this document Project Report).
2. Design a data warehouse using one of the schemas presented in class. Justify your choice of schema. For your schema, identify and justify dimensions and their attributes, fact table and its attributes, concept and schema hierarchy (ies) for each dimension. All of your dimensions, their table's definitions, and justifications along with all of your diagrams (Only the computer-generated diagrams are acceptable) must be included in the document Project Report under proper headings of your choice.
3. Create your data warehouse (as a database) using MYSQL (a free relational database management system, RDBMS, that can be downloaded from internet). Populate your data warehouse using the integrated data. Your mechanism for finding "facts" out of the integrated data for using in the subject table of your data warehouse is of a great interest. Clearly describe how your mechanism of choice does the work.  
Your explanations and justifications about the data warehouse creation and the mechanism for the "fact" findings must be included in the Project Report document under proper headings of your choice.
4. Write an interface for the data warehouse (Computer language of your choice) that mimics an OLAP system. That is, for the cuboid of your data warehouse user can issue a sequence of OLAP operations to express a query against the interface.



The interface has two tasks: (i) help the user to enter the query in OLAP operations by guiding the user and (ii) creates a set of SQL codes that will deliver the output of the OLAP operations. Therefore, as one solution, a template may be created by the interface to guide the user to fill in a template. By doing so, you can capture the necessary information to be used by the SQL

The interface:

- a. Accepts a query expressed by the OLAP operations as input
- b. Converts the query into SQL statements
- c. Submits the SQL statement to MYSQL for implementation and displays the result produced by MYSQL.

5. Test your interface by implementing the following four queries using OLAP operations: (Express each query using OLAP operations and convert the OLAP operations into SQL statements by the help of the interface, and then implement the SQL statements.)
  - a. *Get the number of graduates for the Cyber College for 89*
  - b. *Get the number of graduates with a BS in Computer science in summer of 1988.*
  - c. *Get the number of graduates with a high GPA in the college of business for Spring semesters of 90 and 92.*
  - d. *Get the number of international students graduated in years 89, 90, and 91*

For each one of the above queries, include the followings in the Project Report under proper headings:

- i. Query itself in plain English (as shown above),
  - ii. The query codes using OLAP operations by the help of the Interface (the outcome of this step will be used as input to the interface),
  - iii. List of SQL codes produced by your interface from translation of the output of step ii (above) into SQL, and
  - iv. The outcome produced by MYSQL implementing the output of step iii. (You are not permitted to type this outcome. I must see the outcome as it is generated by MYSQL. You may take a screen shot, print it, and then include the printed image in your Project Report).
6. Add a cover page, table of contents, table of figures, introduction, and a conclusion to your Project Report. Also, add your program codes as an appendix to the Project Report and submit your Project Report by email to me not later than **Nov. 27, 2023**
  7. You must submit your assignment on an incremental basis as follows:
    - A. The deadline for submitting the completed **item 1** (above) along with the Project Report of the completed item 1 to me is **Oct. 2, 2023**. (You can consider this deliverable as the first assignment for your course, and it will be graded accordingly).
    - B. The deadline for submitting the completed **items 2** (above) along with the Project Report of the assignment for the completed items 1 and 2 to me is

**Oct. 13, 2023.** (You can consider this deliverable as your **second** assignment, and it will be graded accordingly.)

- C. The deadline for submitting the completed **item 3** (above) along with the Project Report of the assignment for the completed items 1, 2, and 3 to me is **Oct. 26, 2023**. (You can consider this deliverable as your **Exam 2** and it will be graded accordingly.)
- D. The deadline for submitting the completed **items 4, 5, and 6** (above) along with the Project Report of the assignment for the completed items 1, 2, 3, 4, 5, and 6 to me is **Nov. 27, 2023**. (You can consider this deliverable as a take home test for your **Final Exam**, and it will be graded accordingly.)
- E. Every one of you will have a private zoom meeting with me on Nov. 29 and Dec. 24 to demo your project. During this meeting,
  - I will give you a query in plain English,
  - You will write the query using OLAP operators by the help of the interface to generate the SQL codes and display them,
  - Feed the codes to MYSQL and display the outcome.

If your system fails, then the failure has a great effect on your score for the final exam.

## **Extracted Data**

Extracted data from 3 data sources (Data source1, Data source2, and Data source3) by the university authorities

Extracted from data sources 1 and 2:

COLLEGES	
NAME	MAJORS
Cyber College	Computer Sc Information Sc Applied Sc
College of Business	Accounting Business Admin Economics
College of Education	Elementary Ed Secondary Ed
College of Art and Science	Biology Chemistry English

If you find majores with misspelling or not exact match with the list of majors displayed above, then they need to be fixed as part of the integration process.

## STUDENTS

ID	Name	Address	Major	Degree	GPA	MonthDay	Y	Status
933	GG. Art	Savannah	English	BA	3.1	Aug.15	91	N
941	BB. Mercy	Savannah	English	MA	4	Aug. 15	91	N
356	C. Brown	Denver	Applied Sc	PhD	4.0	Dec.15	89	I
667	K. Chen	Conway	Elementary	EdD	3.2	Aug. 15	92	O
529	M. Yu	Conway	Secondary	MS	3.2	May 15	90	I
552	K. Ahmad	Atlanta	Secondary	EdD	3.9	July 48	87	O
687	Y. Morty	Atlanta	Computer Sc	BS	3.9	July 4	84	N
688	Y. Crema	Little Rock	Biology	AS	2.2	June 1	87	N
653	C. Alshukri	Denver	Applied Sc	PhD	4.0	Dec.15	91	O
100	A. Jones	Little Rock	Computer Sc	BS	3.1	May 15	84	I
986	Q. Brady	Pine Bluff	Computer Sc	MS	3.8	May 15	87	N
579	U. Jones	Little Rock	Computer Sc	BS	3.1	May 15	84	N
200	B. Brady	Little Rock	Applied Sc	PhD	3.2	May 15	92	I
300	C. Cook	Little Rock	Biology	BS	2.2	June 1	87	I
400	D. Morty	Atlanta	Biology	BS	3.9	July 4	84	N
422	U. Yu	Conway	Secondary	MS	3.2	May 15	90	O
423	U. Mooshe	Atlanta	Secondary	EdD	3.9	July 4	90	O
401	D. Morgan	Conway	Biology	BS	3.9	July 4	87	I
702	Y. Brady	Conway	Applied Sc	PhD	3.2	May 15	87	I
201	W. Brady	Conway	Applied Sc	PhD	3.2	May 15	87	I
202	A. Bank	Conway	Applied Sc	PhD	3.2	July 4	87	N
415	T. King	Conway	Elementary	EdD	3.2	Aug. 15	90	N
901	C. Crema	Little Rock	Biology	AS	2.2	June 1	87	O

222	D. Abul	Savannah	Economics	BA	3.9	Aug. 15	90	N
333	G. Halk	Atlanta	B Admin	MBA	2.9	Aug. 15	90	N
431	W.Goos	Savannah	Accounting	BA	4.1	Aug. 15	90	O
421	T. Chen	Conway	Elementary	EdD	3.2	Aug. 15	91	I
524	U. Alramzy	Denver	Applied Sc	PhD	4.0	Dec.15	92	I
689	Y. Morgan	Conway	Biology	BS	3.9	July 4	87	N
195	D. Red	Seattle	Info Sc	BS	2.1	Dec. 15	90	O
703	Y. Bank	Conway	Applied Sc	PhD	3.2	July 15	87	O
665	Z. Silver	Seattle	Elementary	EdD	3.3	Aug. 15	91	I
739	A. Moon	Little Rock	Computer Sc	BS	3.1	June 1	84	N
839	Q. Robe	Pine Bluff	Computer Sc	MS	3.8	June 18	87	N
939	U. Wang	Little Rock	Computer Sc	AS	3.1	July 15	84	O
356	C. Brown	Denver	Applied Sc	PhD	4.0	Dec.15	89	O
500	K.Bronz	Dallas	Chemistry	MS	3.2	July 4	88	O
132	A. White	Orlando	Computer Sc	AS	3.9	Aug 15	92	N
100	A. Jones	Little Rock	Computer Sc	MS	3.2	May 15	87	N
232	K. Blue	Little Rock	English	AS	3.1	Aug 15	91	I
352	C. Black	Pine Bluff	Applied	PhD	3.8	Aug 15	90	O
600	C. Jones	Dallas	Info Sc	BS	4.0	Dec. 15	88	O
680	Y. Brady	Pine Bluff	Computer Sc	MS	3.8	May 15	90	O
682	Y. Cook	Little Rock	Biology	BS	2.2	June 1	87	O
400	D. Morty	Seattle	Info Sc	BS	2.1	Dec. 15	88	I
452	N. Brady	Little Rock	Applied Sc	PhD	3.2	Aug 15	88	I
590	X. Cook	Little Rock	Biology	AS	2.2	June 1	92	N
690	D. Brown	Atlanta	Biology	BS	3.9	July 4	88	N
175	K. Glob	Dallas	Applied Sc	PhD	3.2	July 4	89	N
681	Y. Brady	Little Rock	Applied Sc	PhD	3.2	May 15	91	N
185	X. Jones	New York	Info Sc	BS	4.0	Dec. 15	90	N
700	A. Boss	Seattle	Elementary	MS	3.3	Aug. 15	84	I
800	K. Queen	Conway	Elementary	EdD	3.2	Aug. 15	89	O
416	T. Gola	Conway	Secondary	MS	3.2	May 15	90	N
420	T. Silver	Seattle	Elementary	EdD	3.3	Aug. 15	91	N
900	B. Gola	Conway	Secondary	MS	3.2	May 15	87	I
955	K. Booth	Atlanta	Secondary	MS	3.9	May 15	87	O
956	C. Jessy	Orlando	Applied Sc	MS	4.0	Dec.15	89	O
150	A. Cool	Orlando	Computer Sc	MS	3.9	Aug 15	86	I
250	K. Jones	Little Rock	Computer Sc	BS	3.1	Aug 15	86	N
350	C. Cool	Pine Bluff	Applied	MS	4.0	Aug 15	86	N
229	M. Katchal	Savannah	Economics	BA	3.0	Aug. 15	91	N
339	H.Poory	Atlanta	B Admin	BA	3.9	Aug. 15	87	N
439	H. Mishu	Savannah	Accounting	BA	2.8	Aug. 15	87	O
329	G. Kooper	Savannah	Accounting	BA	3.0	Aug. 15	91	N
339	M. Bayraq	Savannah	MBA	MBA	3.9	Aug. 15	87	N
337	N. Mikhy	Savannah	B Admin	BA	2.8	Aug. 15	87	O
650	D. Sorty	Atlanta	Biology	BS	3.9	July 4	84	N
750	K.Gold	Dallas	Applied Sc	PhD	3.2	July 4	84	N
850	F. Jones	New York	Info Sc	BS	4.0	Dec. 15	85	I
950	D. Golden	Seattle	Info Sc	BS	2.1	Dec. 15	85	I
165	A. Silver	Seattle	Elementary	MS	3.3	Aug. 15	85	O
417	T. Dool	Atlanta	English	MS	3.9	July 4-9	92	O
418	T. Brown	Denver	Applied Sc	PhD	4.0	Dec.15	92	O
419	T. Red	Seattle	Info Sc	BS	2.1	Dec. 15	92	O
215	K. King	Conway	Elementary	EdD	3.2	Aug. 15	84	N

315	M. Gola	Conway	English	MS	3.2	May 15	84	N
255	K. Dool	Atlanta	Engl	MS	3.9	July 4	86	N
299	M. Golab	Seatel	Chemistry	BS	3.0	May 15	92	N
199	K. Duell	Atlanta	Biology	AS	4.0	Dec. 15	88	I
473	M. Browny	Seattle	CS	PhD	4.0	May.15	92	I
407	T. Doolak	Seattle	BusinessAD	MBA	3.1	June 1	92	O
205	K. Minges	Conway	Elementary	EdD	3.2	Aug. 15	84	N
305	H. Molar	Conway	Biology	MS	3.2	July 4	84	N
207	K. Goolar	Atlanta	Chemistry	MS	3.9	July 4	86	N
107	M. Kochaal	Savannah	Chemistry	BS	3.9	May 15	92	N
305	M. Searsy	Conway	Biol	MS	3.2	July 4	84	N
207	P. Little	Atlanta	Chemistry	MS	3.9	July 4	86	N
107	M. Big	Little Rock	Chemistry	BS	3.9	May 15	92	O
109	P. Board	Atlanta	Chem	AS	4.0	Dec. 15	88	I
305	M. Korp	Conway	Biology	MS	3.2	July 4	84	N
109	G. Bear	Little Rock	Biology	AS	4.0	Dec. 15	88	O
591	Z. Kochaal	Savannah	Chemistry	BS	3.9	May 15	92	N
391	X. Searsy	Conway	Biology	MS	3.2	July 4	84	N
201	Y. Little	Atlanta	Chemistry	MS	3.9	July 4	86	N
117	W. Big	Little Rock	Chemistry	BS	3.9	May 15	92	O
612	V. Board	Atlanta	Chemistry	AS	4.0	Dec. 15	88	I
305	M. Kartsoo	Conway	Biology	MS	3.2	July 4	84	N
609	N. Gearsal	Little Rock	Biology	BS	4.0	Dec. 15	88	O
615	K. Jiring	Conway	Element.	BA	3.2	Aug. 15	84	N
715	M. Chendo	Conway	Secondary	MS	3.2	May 15	84	N
855	Z. Megart	Atlanta	Secondary	MS	3.9	July 4	86	N
699	A. Woo	Seattle	Elementary	BA	3.0	May 15	92	N
933	CC. Artian	Atlanta	English	BA	2.9	Aug.15	91	N
941	D. Merdoo	Atlanta	English	MA	2.8	Aug. 15	91	N

### Extracted data from data source 3

COLLEGES		MAJORS
NAME		
Computer Communications (Old name for the Cyber College)		Computer Sc Information Sc Applied Sc
College of Business and IS (Old name for the College of Business)		Accounting Business Admin Economics
College of Education		Elementary Ed Secondary Ed
College of Science (Old name for the College of Art and Science)		Biology Chemistry

ID	Name	Address	Major	Degree	GPA	Semester	Y	Status
999	CK. Brown	Denver	Applied Sc	1	4.0	Sum1	89	1
961	QT. Brady	Pine Bluff	Computer Sc	1	3.8	Sum2	87	1
962	UL. Jones	Little Rock	Computer Sc	1	3.1	Sum2	84	2
970	Bf. Brady	Little Rock	Applied Sc	1	3.2	Sum1	92	2
975	CM. Cook	Little Rock	Biology	2	2.2	F	87	2
412	DP. Morty	Atlanta	Biology	3	3.9	S	84	2
971	UK. Yu	Conway	Secondary	4	3.2	F	90	2
989	Li. Mooshe	Atlanta	Secondary	4	3.9	F	90	3
998	Di. Morgan	Conway	Biology	4	3.9	S	87	3
964	YH. Brady	Conway	Applied Sc	3	3.2	Sum1	87	3
977	WT. Brady	Conway	Applied Sc	3	3.2	Sum2	87	3
987	AC. Bank	Conway	Applied Sc	3	3.2	Sum2	87	3
988	TP. Kim	Conway	Elementary	2	3.2	Sum1	90	3
902	CY. Crema	Little Rock	Biology	2	2.2	S	87	3
969	D. Abuly	Savannah	Economics	2	3.9	S	90	5
968	G. Halky	Atlanta	B Admin	2	2.9	S	90	5
967	W.Goast	Savannah	Accounting	2	4.1	F	90	5
833	T. Chenty	Conway	Elementary	1	3.2	Sum1	91	5
124	U. Algy	Denver	Applied Sc	1	4.0	S	92	1
976	YL. Torgan	Conway	Biology	1	3.9	S	87	1
100	D. Reddyg	Seattle	Info Sc	4	2.1	F	90	1
969	Y. Bankrpt	Conway	Applied Sc	4	3.2	Sum1	87	2
901	UL. Wagon	Little Rock	Computer S	4	3.1	Sum1	84	2
902	C. Browner	Denver	Applied Sc	4	4.0	S	89	2
777	F. Anglic	Atlanta	Biology	2	3.9	S	90	3
779	BB. Goofan	Atlanta	Chemistry	2	2.1	F	91	3



**The following information was extracted from the data dictionaries:**

**In Data sources of 1 and 2**

Fall semester ends on Dec. 15	Spring semester ends on May. 15
Summer 1 ends on June 1	Summer 2 ends on July 4
Summer 3 ends on July 15	Summer 4 ends on Aug 15

Status:	I	International Student
	O	Out of state
	N	Non-of-the above
Degrees	AS	Associate degree
	BS	Bachelor of Science
	BA	Bachelor of Art
	MS	Master of Science
	MA	Master of Art
	MBA	Master of Business Administration
	EdD	Doctor of Education
	PhD	Doctor of Philosophy

**In Data source3**

The name of semesters is used as follows: F (for Fall), S (for Spring), Sum1, Sum2

Status:	1	International Student
	2	Out of state
	3	Non-of-the above
Degrees	1	Associate degree
	2	Bachelor of Science
	3	Bachelor of Art
	4	Master of Science
	5	Doctor of Education

**Extra information:**

The assumption is that students have a unique ID. If you find two students with the same ID, please change one of them to a different 3-digit number.

4 ≥ GPA ≥ 3 is considered High,  
3 > GPA > 2 is considered Medium, and  
2 ≥ GPA is considered Low.