

 اونیورسیتی ملیسیا قهغ UNIVERSITI MALAYSIA PAHANG PUSAT SAINS MATEMATIK	SUBJECT: BSD2343 DATA WAREHOUSING		MARKS: 90(30%)
	GROUP PROJECT	DUE DATE: Group Project Report: 14/6/2022 11.59 PM Group Project Presentation: 15/6/2022	

CLO	Description	PLO mapping	Marks	Percentage
CLO1	Acquire fundamental Big data and data warehousing concepts	PLO1: Able to possess fundamental knowledge of Data Analytics. C3: Knowledge and understanding	20	6.67%
CLO2	Analyze real life problems using appropriate Big data and data warehousing concepts	PLO2: Able to develop analytical and critical thinking by utilizing Data Analytics knowledge in solving various problems C4: Critical thinking and problem solving.	20	6.67%
CLO3	Build and integrate Big data in data warehouse by using appropriate software	PLO3: Modern tool usage. P4: Mechanism.	30	10%
CLO4	Work in group in order to complete the given assessments in specific time frame.	PLO4: Individual & Teamwork A3: Valuing	20	6.67%

Summary

The worldwide data warehousing market is expected to be worth more than \$30 billion by 2025. Data warehousing and analytics will play a significant role in a company's future growth and profitability. Data warehouse solutions will provide every business with considerable advantage by evaluating all the data they collect and making better decisions. Understanding business data will help in making intelligent business decisions that determine whether an organization succeeds or fails. The demand for Big Data and Data Analytics will continue to grow in the coming days, leading to a greater need for Data Warehouse solutions.

To make real progress along the path toward becoming a data warehouse engineer, it is useful to apply all your data warehousing knowledge to a project task. This project will require you to explore the real data and gain experience in implementing the data warehousing project and become industry ready.

Requirements:

For this group project, you are required to:

- form a group consisting of **FOUR (4) OR FIVE (5)** members.
- browse any datasets with huge volumes (big data) that are available online through Kaggle/GitHub or any repository. Your datasets must contain at least 5 tables and cannot be processed using traditional technologies, processing methods, algorithms, or any commercial off-the-shelf solutions. This dataset will be the title of your project and your main problem to be solved. For example, if you choose the E-commerce dataset, then, your project title will be 'Solving Problem in E-Commerce with Data'. Make sure your problem differs from other groups.
- utilize any data warehouse and big data tools that you have learned in this course to solve the problem.
- Please include all the required information as stated in the general instructions below.

General Instructions:

Submission of group project report should require all the followings:

1. Background

Description of the selected project, the problem to be solved, and data schema.

2. Architecture

Sketch the pipeline structure of your project and provide a description of the process flow.

3. Database

Present your database in a relational model and identify the relationship between data. Then, map the relationship between data by using appropriate tools. Identify your data warehouse schema.

4. ETL Pipeline

To load data from your OLTP (Database) into OLAP (Data Warehouse), you need a pipeline for ETL. To do that, you need to create an ETL process for each dimension and fact table.

5. Results and Data Analysis

After your data is clean from the ETL process, now its time to perform some analysis such as roll-up, slicing, pivot and others. Visualize your analysis through appropriate data visualization tools and techniques.

6. Conclusion

Conclude your project and discuss all the challenges that you face in this project.

7. References

Cite every single reference used in completing the project.

For the cover page, please make sure to write your group name, name of members, matric number and section on ONE page. Submit your assignment (pdf files) to the KALAM platform latest by the due date, 14th June 2022. Submission after the due date is considered as a late submission and the mark will be deducted by 5%. The project presentation will be held on Week 14.

Plagiarism:

Your attention is drawn to the no-plagiarism policy. This covers cheating, plagiarism, collusion and any other attempts to gain an unfair advantage in assessment. The work you submit must conform to this policy. If plagiarism is discovered, any parties involved will take equal blame (will get zero mark).

BSD2343 DATA WAREHOUSING GROUP PROJECT MARKING SCHEME

Rubric for CL01

CLO1: Acquire fundamental Big data and data warehousing concepts								
CRITERIA	LEVEL OF ACHIEVEMENT						WEIGHT AGE	SCORE
	0 Grossly Inadequate	1 Inadequate	2 Emerging	3 Developing	4 Good	5 Excellent		
Description and explanation of the selected project and problem to be solved.	No description and explanation about the project selected and problem to be solved in the report.	Poorly describe and explain about the project selected with no problem to be solved in the report.	Poorly describe and explain about the project selected and problem to be solved in the report.	Fairly describe and explain about the project selected and problem to be solved in the report.	Clearly describe and explain about the project selected and problem to be solved in the report.	Excellently describe and explain about the project selected and problem to be solved in the report.	1	
Explanation regarding the data schema and the relationship between data.	Failed to explain the data schema and the relationship between data.	Able to explain the data schema but failed to explain the relationship between data.	Poorly explain the data schema and the relationship between data.	Fairly explain the data schema and the relationship between data.	Good explanation on the data schema and the relationship between data.	Excellent explanation on the data schema and the relationship between data.	2	
Concluding remarks.	No concluding remarks provided.	Limited concluding remarks provided and inaccurate.	Concluding remarks provided but unclear and inaccurate.	Concluding remarks provided but partly inaccurate.	Clear and good concluding remarks provided.	Very clear and excellent concluding remarks provided.	1	
						TOTAL		20

Rubric for CL02

CLO2: Analyze real life problems using appropriate Big data and data warehousing concepts								
CRITERIA	LEVEL OF ACHIEVEMENT						WEIGHT AGE	SCORE
	0 Grossly Inadequate	1 Inadequate	2 Emerging	3 Developing	4 Good	5 Excellent		
Able to sketch the pipeline structure of the project.	Failed to provide the pipeline structure of the project.	Able to sketch the pipeline structure but the pipeline is wrong.	Able to sketch the pipeline structure but the pipeline is partly correct.	Able to sketch the pipeline structure correctly but no description on the pipeline.	Able to sketch the pipeline structure correctly but limited description on the pipeline.	Excellently sketch the pipeline structure with good description on the pipeline.	1	
Able to present database in relational model and map the relationship between data.	Failed to present database in relational model and map the relationship between data.	Able to present database in relational model but failed to map the relationship between data.	Poorly present database in relational model and map the relationship between data..	Fairly present database in relational model and map the relationship between data.	Clearly present database in relational model and map the relationship between data.	Excellently present database in relational model and map the relationship between data.	1	
Able to sketch and explain the ETL pipeline.	Failed to sketch and explain the ETL pipeline.	Able to sketch ETL pipeline but failed to provide any explanation on the pipeline.	Poorly sketch and explain the ETL pipeline.	Fairly sketch and explain the ETL pipeline.	Clearly sketch and explain the ETL pipeline.	Excellently sketch and explain the ETL pipeline.	1	
Perform all data analysis with data visualization.	No results obtained from the data analysis.	Poorly present data analysis and there is no data visualization.	Poorly present data analysis and data visualization.	Fairly present data analysis and data visualization.	Clear and good presentation of the data analysis and data visualization.	Excellent presentation of the data analysis and data visualization.	0.5	
Explanation on the data analysis results.	Unable to provide any explanation on the data analysis results.	Limited explanation on the data analysis results.	Able to explain the data analysis results but unclear and inaccurate.	Able to explain the data analysis results clearly but inaccurate.	Able to explain the data analysis results clearly.	Able to explain the data analysis results perfectly.	0.5	
						TOTAL	20	

Rubric for CLO3

CLO3: Build and integrate Big data in data warehouse by using appropriate software								
CRITERIA	LEVEL OF ACHIEVEMENT						E	W
	0 Grossly Inadequate	1 Inadequate	2 Emerging	3 Developing	4 Good	5 Excellent		S
Ability to extract the datasets from sources very well.	Unable to extract the datasets.	Barely able to extract the datasets from the sources.	Partly able to extract the datasets from the sources.	Able to extract the datasets from the sources in successful results.	Very good in extracting the datasets from the sources.	Able to extract the datasets from the sources excellently.	1	
Ability to construct pipeline structure of the project involving various tools from Big Data and Data Warehouse.	Unable to construct pipeline structure of the project involving various tools from Big Data and Data Warehouse.	Barely able to construct pipeline structure of the project involving various tools from Big Data and Data Warehouse.	Able to construct pipeline structure of the project but limited tools from Big Data and Data Warehouse.	Able to construct pipeline structure of the project involving various tools from Big Data and Data Warehouse in successful results.	Very good in constructing pipeline structure of the project involving various tools from Big Data and Data Warehouse.	Able to construct pipeline structure of the project involving various tools from Big Data and Data Warehouse excellently.	1	
Able to map the relationship between data by using appropriate tools.	Unable to map the relationship between data by using appropriate tools.	Barely able to map the relationship between data by using appropriate tools.	Able to map the relationship between data by using appropriate tools but the mapping is inaccurate.	Able to map the relationship between data by using appropriate tools correctly.	Very good in mapping map the relationship between data by using appropriate tools.	Able to map the relationship between data by using appropriate tools excellently.	1	
Ability to construct ETL pipeline by using appropriate tools.	Unable to construct ETL pipeline by using appropriate tools.	Barely able to construct ETL pipeline by using appropriate tools.	Able to construct ETL pipeline by using appropriate tools but the pipeline is inaccurate.	Able to construct ETL pipeline by using appropriate tools correctly.	Very good in constructing ETL pipeline by using appropriate tools.	Able to construct ETL pipeline by using appropriate tools excellently.	1	
Ability to visualise (table, graph, GUI and etc) the programming codes.	Unable to visualise (table, graph, GUI and etc) the programming codes.	Barely able to visualise (table, graph, GUI and etc) the programming codes.	Partly to visualise (table, graph, GUI and etc) the programming codes.	Ability to visualise (table, graph, GUI and etc) the programming codes in successful results.	Very good in visualising (table, graph, GUI and etc) the programming codes.	Excellent visualise (table, graph, GUI and etc) the programming codes.	1	
The code can be executed and easy to understand the codes constructed.	No code is constructed.	Only few codes can be executed and difficult to follow the structure and flow of the codes.	Some of the codes can be executed and fairly difficult to follow the structure and flow of the codes.	The code can be executed and fairly easy to follow the structure and flow of the codes.	The code can be executed and easily to follow the structure and flow of the codes.	The code can be executed and well easily to follow the structure and flow of the codes.	1	
						TOTAL		30

Rubric for CLO4

CLO4: Work in group in order to complete the given assessments in specific time frame.								
CRITERIA	LEVEL OF ACHIEVEMENT						WEIGHT AGE	SCORE
	0 Grossly Inadequate	1 Inadequate	2 Emerging	3 Developing	4 Good	5 Excellent		
Every member in the group able to provide information related to this project and show understanding of the project.	All group members unable to provide information related to this project and show understanding of the project.	Some group members unable to provide information related to the project and show understanding of the project.	One or two group members unable to provide information related to the project and show understanding of the topic.	All group members able to provide adequate information related to this project but only show average understanding of the project.	All group members able to provide adequate information related to this project and show good understanding of the project.	All group members able to provide information related to this project and show excellent understanding of the project.	1	
Able to work together as a team towards goal achievement and submit report on time.	Unable to work together as a team and submit report on time.	Able to work together as a team, however, unable to submit report on time.	Able to work together as a team towards goal achievement, however, unable to submit quality report.	Able to work together as a team towards goal achievement and submit average quality report on time.	Able to work together as a team towards goal achievement and submit good quality report on time.	Able to work together as a team towards goal achievement and submit high quality report on time.	2	
Ability to assume alternate roles as a group leader and group members.	No clear evidence of ability to assume alternate roles as a group leader and group members demonstrated in practice.	Attempt to demonstrate in practice the ability to alternate roles as a group leader and group members but with limited effect and require improvements.	Able to demonstrate in practice the ability to assume alternate roles as a group leader and group members with some effect(s) and require minor improvements.	Able to demonstrate in practice the ability to assume alternate roles as a group leader and a group member to achieve the same goal.	Able to demonstrate in good practice the ability to assume alternate roles as a group leader and a group member to achieve the same goal.	Able to demonstrate in excellent practice the ability to assume alternate roles as a group leader and a group member to achieve the same goal.	1	
						TOTAL		20

