Subject Code	**WQD 7009**
Subject Name	**Big Data Applications and Analytics**
Course Leader	Dr. Riyaz Ahamed
Weightage	40 %
Assignment	Group Assignment
Handout date	15th November 2024
Submission date	27th December 2024
Instructions	1. This is a group assignment carrying 40% of the total marks available for the module. Each group should consist of **a maximum of 5 members only.** 2. There is no guarantee that all the members in a group will be getting identical marks, because members will be assessed individually based on different criteria, which include Q & A session, an understanding of the whole concept of the assignment. Any free riders shall be penalized with zero marks. 3. The output of this assignment is in terms of **data processing and analytics architecture, visualization, comparison, evaluation, and presentation.** 4. Please remember to structure your presentation to fit within the allotted 10 to 15 minutes. Effective time management is crucial and will impact on your marks. **Zero marks will be assigned if any group member misses the presentation or demo session.** 5. The output of this assignment is in terms of a report written within the range of between 3000 to 3200 words. 6. You are required to submit your report in softcopy (through Spectrum). Kindly ensure name and ID is written on the cover sheet. GENTLE REMINDER: 7. Plagiarism is a serious offence and plagiarized work will result in an f grade. Failure to submit the report by the deadline shall be penalized with zero marks.

Task description

In this assignment, you are required to: -

a) Brainstorm and select the best dataset from your individual assignment based on the climate change and sustainability dataset for the group assignment. Attach evidence of the dataset selection from your individual assignment to carry forward in the group assignment. Include meeting minutes discussing the dataset selection, considering features like relevance to objective, data quality, size of the dataset, data variety, relevance to industry/domain, data exploration potential, compatibility with tools, potential for innovation, and interoperability.

b) For the chosen dataset, propose a cloud-based data analytics framework for your topic. The framework should cover various layers of the data lifecycle process, including data ingestion, storage, processing, analytics, visualization, and other layers based on your project's requirements. Refer to class discussions and research papers for different layers in your data lifecycle or data architecture.

c) Select and justify one potential tool relevant to each layer of the data lifecycle or data architecture process as proposed in (b). The chosen tools can be either on-premises tools such as Spark, MapReduce, Hive, PowerBI, etc., or cloud-based tools from Google Cloud, AWS, Azure, or any other cloud platform. You have the option to include both on-premises and cloud tools for your framework development. Each layer should have one tool and its justification.

d) Produce practical output evidence by implementing at least 30% of the proposed data lifecycle process in (b). 30% should cover at least two tools in your framework development. It’s recommended to implement a cloud platform for the data processing and analytics framework.

e) Evaluate the implemented framework from part (d) using three (3) key performance metrics related to query processing and data engineering. Metrics may include processing speed, memory consumption, compatibility, data pipeline efficiency, data throughput, scalability, or fault tolerance. Present your analysis with three supporting graphs to effectively demonstrate the framework’s performance and its relevance in data engineering contexts.

The following deliverables must be completed for the chosen topic.

Assessment Marking Criteria - Total: 80 Marks (Normalized to 40%)

1. Introduction – 5 Marks
2. Meeting minutes report – 5 Marks
3. Different layers in data lifecycle process or data architecture diagram and its explanation. (Explain how it works in the architecture) – 15 Marks.
4. On premises or cloud-based tools selection and its justification – 10 Marks.
5. Implementation of the proposed framework (Practical) – 25 Marks
6. Evaluation metrics with graph -10 Marks
7. Presentation & demonstration – 10 Marks

MARKING RUBRICS

Category	Highly Competent	Competent	Satisfactory	Unsatisfactory
Introduction	A well-articulated introduction that provides a clear, logical, and succinct description of content, scope, and organization. Current problems and objectives were addressed well. **(5 Marks)**	An introduction that describes the content, scope, and organization of the review. Current problems and objectives were clearly related to the topic. **(4-3 Marks)**	An introduction that outlines the content, scope, and organization of the review. There is some indication of the problem addressed and objectives for the topic. **(2 Marks)**	Either no introduction or one that poorly or partially situates the reader in the context of the concern. No evidence for the problem statement and objectives in the document. **(1 Mark)**
Meeting Minutes Report	Comprehensive and well-documented meeting minutes that effectively capture the discussion around dataset selection, considering relevant features. **(5 Marks)**	Meeting minutes report that adequately captures the discussion around dataset selection and relevant features. **(4-3 Marks)**	Meeting minutes report that provides basic details of the discussion around dataset selection, with some relevant features mentioned. **(2 Marks)**	Incomplete or missing meeting minutes report related to dataset selection and relevant features. **(1 Mark)**
Five layers data life cycle process diagram and its explanation. –	Good explanation of the various components of the framework. Excellent diagram that clearly shows all the components of the framework and their relationship and communication between them. **(15-12 Marks)**	Moderately explained the components of the framework. Clear diagrams show all the components and the relationship between them. **(11-07 Marks)**	Made an attempt to draw and explain the components of the framework. The report includes the development of a few operational definitions for key terms and/or there are few or no connections among key terms. **(06-03 Marks)**	No diagram and minimum explanation of the proposed framework. **(02-01 Marks)**

Tools selection justification	Good explanation and justification for the selected tools. **(10-8 Marks)**	Moderately explained and justified the selected tools for the proposed framework. **(07-05 Marks)**	Made an attempt to provide justification for the selected tools for the proposed framework. **(04-03 Marks)**	No or minimum justification for the selected tools of the proposed framework. **(02-01 Marks)**
Framework implementation	A minimum of three tools should be implemented. Attached evidence for 30% of the data life cycle process. The report has a well-defined conceptual framework that includes all the five layers of the framework and clearly articulates relationships among key terms. Clearly identifies a range of possible alternative technologies to create an understanding of the product and chooses the most appropriate to perform the task. **(25 - 18 Marks)**	A minimum of two different tools should be implemented. The final report includes development operational definitions for most key terms that are loosely linked into an overall conceptual framework. Provided clearer relationships among all the five layers of the framework. Applies selected technology to perform the task with some efficiency and effectiveness. Displays understanding of the results. **(17-11 Marks)**	The final report includes the development of a few operational definitions for key terms and/or there are few or no connections among key terms. Uses selected technology inaccurately to perform tasks ineffectively and inefficiently, while demonstrating only a minimal understanding of the purpose and results. Partial implementation of two different tools. **(10-5 Marks)**	No diagram and minimum explanation of the proposed framework. **(4-1 Marks)**
Evaluation metrics	Clear explanation and analysis of the proposed framework performance. **(10-8 Marks)**	Moderately explained and analyzed the proposed framework performance. **(7-5 Marks)**	Made an attempt to do evaluation and analyze the proposed framework performance. **(4-3 Marks)**	No analysis of the proposed framework performance. **(2-1 Marks)**

End of the Paper