INSTITUTE OF INNOVATION

Java Programming

- Source: Link

Data Analytics

Answer the following questions (10 Marks)

Submission Format (PDF)

1. Define ETL and explain its importance in data management.
2. Describe a scenario where ETL could be beneficial in a business setting.
3. What challenges might a data analyst face during the transformation phase of ETL and how can they be addressed?
4. Explain the concept of data warehousing and its relationship with ETL processes.
5. Define a database and a data warehouse.
6. How do the purposes of a database and a data warehouse differ in a business environment?
7. Can you illustrate with an example when you would use a database versus a data warehouse?
8. List 5 Popular Data Warehouse, ETL Tools and Database.
9. Who is Data Analyst, Business Analyst and Data scientist?
10. Illustrate with an example how data visualization can assist in business decision-making .

Practical (10 Marks)

11. Download a dataset of your choice and build a basic ETL pipeline using Pentaho Data Integration (PDI).

Instructions:

1. Data set Selection:
● Choose a dataset that interests you from a public data repository (e.g.,

Kaggle, UCI Machine Learning Repository, or any government data portal). Ensure the dataset has at least two distinct types of data columns (e.g., numerical and categorical data).

2. Task Overview:
● Describe the dataset you have chosen, including the source, the nature of

the data, and why you selected this particular dataset.

3. ETL Pipeline Design:

● Extract: Outline the steps you took to extract the dataset into Pentaho. Detail any challenges you faced during the extraction phase.

● Transform: Describe the transformations you applied to the data. This could include filtering rows, converting data types, handling missing values, or merging data from multiple sources. Explain the rationale behind each transformation.

● Load: Discuss how you loaded the data into a target system (this could be another database, a flat file, or a data warehouse schema). Describe the structure of the target system and why it was chosen.

4. Challenges and Solutions:

● Highlight any specific challenges you encountered while setting up your ETL pipeline in Pentaho. Discuss how you resolved these issues.

5. Outcome and Reflection:

● Reflect on the process of building the ETL pipeline. Discuss what you learned and how the exercise helped you understand ETL concepts better.

● Optionally, include screenshots of your Pentaho workflow to supplement your explanation.

Submission Requirements:

● A written report that covers all the above points.

● A link to the dataset or a citation of its source.

● Provide the actual Pentaho transformation file (.ktr) as part of your submission.

● Attach Pipeline Images(screenshot) and Output in details along with configuration in Report

● Report can be in word or ppt format

Plagiarism policy:

Plagiarism is an act of copying or including in one's work, work of another (published or unpublished), without adequate acknowledgement, intentionally or unintentionally, for one's benefit. This could include copying another person's work such as the creative expressions, ideas, words, data, figures, assertions, verbatim (cut copy paste) or summarizing/paraphrasing without giving credit to the original source through proper references and/or citations. However, copying the work of another student enrolled in the same module without his/her knowledge shall not be considered under plagiarism, but under malpractice for the purposes of this policy. Submitting one's own work for an assessment, either whole or in part, which is previously submitted for any other course, degree or qualification at this or any other institution, without proper citations and references, is also considered as an act of plagiarism.

It should be your own words, using of AI such as ChatGPT is not allowed and it can be detectible by Turnitin as well as similarity from different resources.

Data Structure

Linked List (25 marks)

- Implement the following functions for a singly linked list:
  - Inserts an element at the specified index.
  - Deletes the element at the specified index.
  - Returns the size of the linked list.
  - Returns true if the linked list is empty, false otherwise.
  - Rotates the linked list by k positions to the right.
  - Reverses the linked list.
  - Appends an element to the end of the linked list.
  - Prepends an element to the beginning of the linked list.
  - Merges two linked lists into a single linked list.
  - Interleaves two linked lists into a single linked list.
  - Returns the middle element of the linked list.
  - Returns the index of the first occurrence of the specified element in the linked list, or -1 if the element is not found.
  - Splits the linked list into two linked lists at the specified index.

Dynamic Arrays (15 marks)

- Implement the following methods for a dynamic array:
  - Inserts an element at the specified index.
  - Deletes the element at the specified index.
  - Returns the size of the dynamic array.
  - Returns true if the dynamic array is empty, false otherwise.
  - Rotates the dynamic array by k positions to the right.
  - Reverses the dynamic array.
  - Appends an element to the end of the dynamic array.
  - Prepends an element to the beginning of the dynamic array.
  - Merges two dynamic arrays into a single dynamic array.
  - Interleaves two dynamic arrays into a single dynamic array.
  - Returns the middle element of the dynamic array.
  - Returns the index of the first occurrence of the specified element in the dynamic array, or -1 if the element is not found.
  - Splits the dynamic array into two dynamic arrays at the specified index.
  - Resizing the arrays with custom factor given by user (other than 2)

Comparison of Linked Lists and Dynamic Arrays (10 marks)

- Compare the following aspects of linked lists and dynamic arrays:
  - Time complexity of each method
  - Space complexity of each method
  - Advantages and disadvantages of each data structure
- Create a report on comparison of both of them