

UNIVERSITY OF BIRMINGHAM

School of Computer Science

Data Structures, Algorithms & Databases

Main Spring Examinations 2023

Data Structures, Algorithms & Databases

Answer ALL questions. Each question will be marked out of 20. The paper will be marked out of 80, which will be rescaled to a mark out of 100.

Question 1 Analysis of Algorithms and Tree Data Structures

Part 1 Consider a function named `arrayToBST` that takes an array of integers a and inserts the elements of a in an **AVL tree** T . Therefore, function `insert` **has re-balancing**.

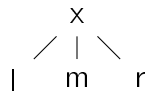
```
Tree arrayToBST(int [] a) {  
    Tree T = □;  
    for (int i = 0; i < a.length; i++) {  
        T = insert(T, a[i]);  
    }  
    return T;  
}
```

Is it possible to construct an input array a of any arbitrary length n that results in the outcomes described below? When it is possible to construct an input, give an example with length $n = 8$.

- (a) Algorithm `arrayToBST` runs in time $\mathcal{O}(n)$. **[4 marks]**
- (b) Algorithm `arrayToBST` runs in time $\mathcal{O}(n \log n)$. **[4 marks]**
- (c) Algorithm `arrayToBST` returns a perfectly balanced tree. **[4 marks]**

Briefly explain your construction. Note that your construction should apply to arrays with any arbitrary length n ; you cannot choose the length. Your input array may or may not have repetitions.

Part 2 Consider a tree data structure where every node has at most three children: left child l , middle child m and right child r . We call it a **ternary tree**.



(d) Give the maximum possible number of nodes of a ternary tree of height 4? **[4 marks]**

(e) Give the minimum possible number of nodes of a ternary tree of height 4? **[4 marks]**

Briefly explain how you compute these numbers. In particular, be clear about the definition of height you are using.

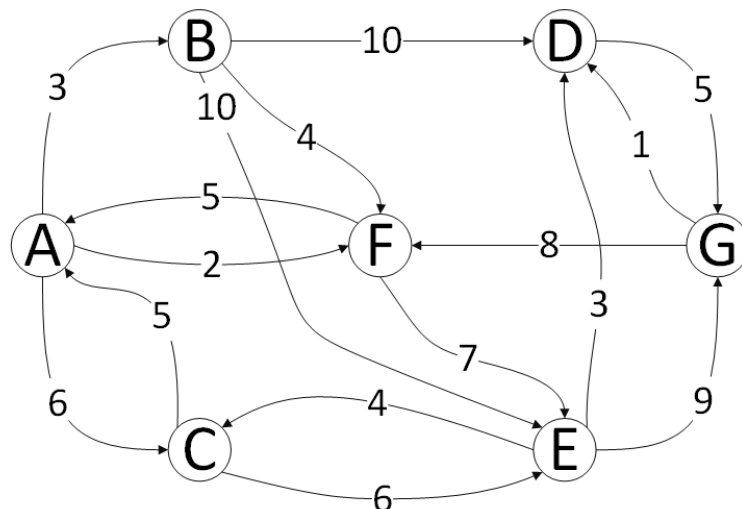
Question 2 Entity-Relationship modelling

A company operates an online store where customers can place orders for various products, and has the following description of their requirements for a database: Customers need to provide details such as their name, email address, and shipping address to create an online account. Each customer may have multiple orders, and each order is associated with only one customer. An order contains one or more order items, where each order item represents a single product ordered by the customer. An order item includes the quantity of the product ordered and its price at the time of purchase. Each product has a unique product ID, a name, and a description. Additionally, the company may track reviews for each product. A review includes the name of the reviewer, a rating, and a comment. Each review is associated with only one product, and each product may have multiple reviews. Finally, each order is associated with a payment. Each payment has a payment ID, an amount, a date and a payment method.

- (a) Develop an **Entity-Relationship model** for the application. Explain the key design decisions made in your choice of entities, relationships and any other (ownership or hierarchical) aspects. **[8 marks]**
Annotate it with **multiplicities** and **hierarchy** annotations, as required. **[4 marks]**
- (b) Carry out **logical design** for the model, representing the design with relational schemas for tables. **[4 marks]**
- (c) Write SQL CREATE TABLE statements for 2–3 tables. Include among them a table that represents a relationship or incorporates a relationship, a weak entity or a subclass entity. The other tables can be for the adjoining entities. **[4 marks]**

Question 3 Graphs and Max-Heap Trees

Part 1 Consider the following weighted directed graph (with 7 vertices and 16 edges):



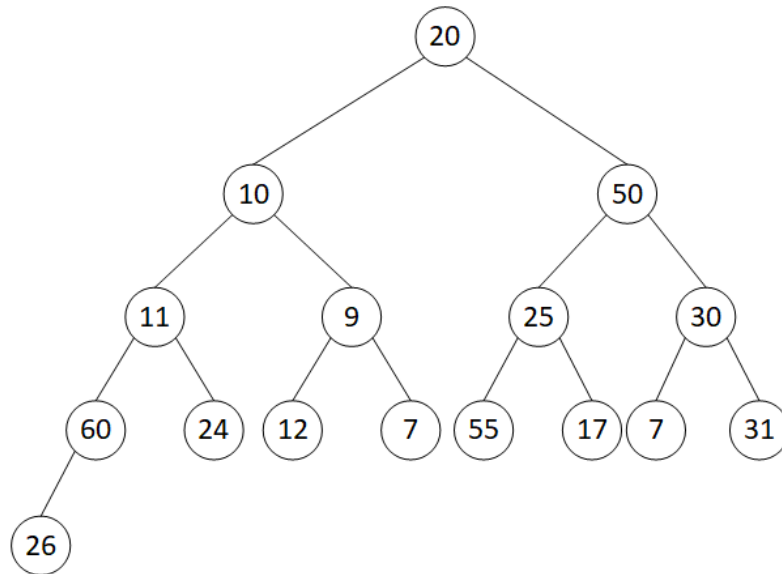
- (a) Calculate the **shortest path** from A to G using the Dijkstra's algorithm. ("Shortest" means the path with the lowest total weight.) **[10 marks]**

You are expected to show your work using a table of the following form and also list the shortest path (e.g. A → B → C) and specify the resulting weight:

A	B	C	D	E	F	G	Finished
0, A	∞ , B	∞ , C	∞ , D	∞ , E	∞ , F	∞ , G	

Total Weight:
Shortest Path:

Part 2 Consider the complete binary tree below. Answer the following questions.



- (b) In the process of building a **Max-Heap Tree**, the first swap is between __ and __ (write down the two numbers). **[2 marks]**
- (c) And the second swap is between __ and __ (write down the two numbers). **[2 marks]**
- (d) Draw the finished Max-Heap Tree (intermediate work is not required). **[6 marks]**

Question 4 Relational algebra and normalisation

Consider the following schema for product sales for an online shop.

product(pid, cat, price)
customer(cid, cname)
sale(pid, cid, date, qty)

(“cat” is short for category and “qty” for quantity). For estimating efficiency, assume around 100 products, 10,000 customers and several million sale records.

- (a) State in English what the following relational algebra expression computes:

$$\pi_{\text{cname}}(\sigma_{\text{qty} * \text{price} \geq 100}(\sigma_{\text{cat} = \text{'clothes'}}((\text{product} \bowtie \text{customer}) \bowtie \text{sale})))$$

[2 marks]

- (b) The expression given above is considered inefficient. State all the ways in which it is inefficient.

[4 marks]

- (c) Give a more efficient version of the expression. Make it as efficient as you can, and explain how it achieves efficiency.

[6 marks]

- (d) Consider a schema $T(A, B, C, D)$ with the functional dependencies $AB \rightarrow C$, $AB \rightarrow D$ and $BC \rightarrow D$.

If we decompose the schema into (A, B, D) & (B, C, D) , is it a **lossless decomposition**? If yes, explain why. If not, try to produce a counterexample.

Are the two subschemas in **Boyce-Codd normal form**? Explain.

[8 marks]