**CSDS 234 Structured and Unstructured Data  
Department of Computer & Data Sciences, Fall 2024**

**Assignment 3**

1. **[Query cost] [25]** Given two tables r and s, where r contains **nr** tuples, s contains **ns** tuples, and **ns>nr>0**. Give the maximum and minimum possible number of tuples for the result produced by the following relational algebra expressions. For each case, also state the assumptions of the schema of r (denoted as **R**) and the schema of s (denoted as **S**) needed to make the expression and results meaningful.
2. r ∪ s
3. r ∩ s
4. r - s
5. r × s
6. σA=2(r)
7. **[Storage] [10]** You are given a hard disk with in total 6 platters. Each platter is double-sided (i.e., has two surfaces that can store data). Each surface has 2K tracks. Each track has 50 sectors. Each sector can store 512 bytes data. Answer the following questions.   
   1. What is the capacity of a single track (the size of the data it can store, in bytes)? A single platter? The entire hard disk?
   2. Identify the valid single block sizes from the following numbers and explain why the rest are not valid: *280 bytes, 1540 bytes, 2048 bytes, 26112 bytes.*
8. **[B+ tree Index] [15]** Consider the following B+ tree (partly shown). Leaf nodes are linked (not shown in the figure). The nodes are annotated with a name (e.g., “I1” for intermediate node 1 and “L2” for a leaf node 2). Answering the following questions.

Diagram

Description automatically generated

1. Give the names of all the tree nodes that must be fetched to answer the query: “find all records with search key greater than 51.”
2. Give the names of all the tree nodes that must be fetched to answer the query: “find all records with search key at least 38 and no larger than 90.”
3. Subtree B is not shown. Nonetheless, what can you infer about this subtree's contents and shape? Put down as much as you can.
4. **[Query Processing -Join] [25]** Let the schema of a relation r as R(A,B,C), and a relation s has schema S(C,D,E). Relation table r has 40K tuples, relation s has 60K tuples. The block factor of r is 25. The block factor of s is 30. Let the average block transfer time is tT. Assume you have a memory that contains M data blocks, but M< 40K/25. Assume there is no index. Estimate the worst-case costs, in terms of the total number of data blocks to be transferred, for the following query r  s using each of the following join strategies:
   * + Nested Loop Join
     + Block Nested Loop Join
     + Hash join
5. **[XML] [15] XML/DTD] [25]** Consider the following XML for Solar System.

<solar\_system>

<star>

<name>Sun</name>

<spectral\_type>G2</spectral\_type>

<age unit="billions years">5</age>

</star>

<planet type="telluric">

<name>Earth</name>

<distance unit="km">149600000</distance>

<mass unit="kg">5.98e24</mass>

<diameter unit="km">12756</diameter>

<satellite number="1"/>

</planet>

</solar\_system>

Add a few lines to the above XML code to include all the facts below about Saturn and Uranus. Your XML needs to be well-formed.

*Both Saturn and Uranus are planets in Solar system. Both Saturn and Uranus have rings. Saturn and Uranus both have type “gaseous”. Saturn is a planet of distance 5.2 (in unit “UA” – astronomical unit), with an Earth mass 95 (unit = “Earth mass”), and a diameter 9.4 Earth diameter. It has 18 satellites. Uranus is a planet of distance 19.2 UA, 14.5 Earth mass, and a diameter of 4 Earth diameter. It has in total 15 satellites.*