ComS 363 Fall 2022
Class Participation Week 11
Topic: Simple query execution plans and cost of full table scan

Learning objectives:

- Practice writing query execution plans with full table scan for SQL queries.
- Practice estimating the cost of a full table scan.

Conceptually, a query execution plan consists of a sequence of relational algebra operators where each operator is associated with the name of the access path (the method/algorithm to retrieve/process the data) for implementing that operator.

Relational operators used in this class participation:

Selection (σ_c) selects a subset of rows from its input where the output rows satisfy the condition c. If the condition is not given, all the input rows are output. Example: $\sigma_{tid>10}(tweets)$ outputs only the rows with tid>10 from the tweets relation.
Projection ($\pi_{ai,\dots}$) outputs attribute values of the specified attributes ai from its input. At least one attribute is required. Duplicate values are eliminated. Example: $\pi_{tid}(tweets)$ outputs only the values of the tid attribute of the tweets table.

Access path used in this class participation:

Assume there are only algorithms: full table scan and on-the-fly. We will learn more algorithms later.

- Full table scan reads one page at a time from a file stored on disk into memory until all the pages that have the data for that relation are read.
- On-the-fly takes an input, performs the task associated with it in memory, and passes the output to the next operator.

Instruction: Submit your attempt to answer the question and sub-questions.

Question: Consider Suppliers(sid int, sname VARCHAR(30) unique not null, address VARCHAR(66), primary key(sid)). Draw a query execution plan for each query below and estimate the cost of each query in terms of the number of disk I/Os using the following assumptions.

Assumption: The Suppliers relation has 10,000 tuples. All Suppliers records are stored in one file using the heap file format. A page/block size is 2,048 bytes where only 2,000 bytes are used for storing tuples. Assume that the fixed-length record format is used to store records in a page. The unpacked bitmap page format is used to store records in a page. Assume that as many records are stored in a page as possible. Assume that one integer takes 4 bytes and one char takes 1 byte. Suppose the selectivity factor of the where clause for 1.c) is 10%. Somesailors is a relation that has only two attributes sid and sname with the same data types as those in the Suppliers table. Assume that this table is empty.

a) select sid, sname from suppliers;
b) select sid from suppliers where sid=10;
c) insert into somesailors
   select sid, sname from suppliers where sid<100;