



CS660 Fall 2023 – Written Assignment 3

Title: *Relational Algebra, Bitmap Index*

Due: 11/03 11:59 PM on Gradescope

Problem 1: Relational Algebra. [35pts]

Consider the following database schema that is part of a larger database and stores information about electronic products (that in our case can be either PC, Laptop or Printer.)

Product (maker, model, type)

PC (model, speed, ram, hdd, price)

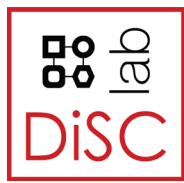
Laptop (model, speed, ram, hdd, screen, price)

Printer (model, color, type, price)

The Product stores the maker, the model and the type that can be either PC, Laptop, or Printer. Each of the other three relations store specific information about each product. Maker and model, or model and type are unique.

Give an expression in relational algebra for each of the queries below:

1. What PC models have speed at least 3?
2. Which manufacturers make laptops with hard disks at least 100GB?
3. Find the manufacturers that make Laptops but not PCs.
4. Find the hard disk sizes that occur in two or more PCs.
5. Find the manufacturer(s) of the PC(s) with the highest available speed.



Problem 2: Relational Algebra. [28pts]

Consider a database with the following schema:

Employee (SSN, name, salary, DNo)

Department (DNo, DeptName, MgrSSN)

Project (PNo, location, ProjName)

HourLog (SSN, PNo, hours)

The Employee relation provides a list of employees with their SSN, name, salary, and department number (DNo). The SSN is unique for each employee. Each employee belongs to only one department. The Department relation contains a list of the departments for the company. Its schema includes a unique department number called DNo. It also includes the name of the department (DeptName) and the social security number of the department's manager (MgrSSN). Each department has only one manager. The Project relation includes a unique project number (PNo), location and the project name (ProjName). An employee can be assigned to any number (including zero) projects. Each project has at least one person assigned to it. Finally, the HourLog relation lists for each project the number of hours of work for each employee who is assigned to that project. The key of this relation is SSN and PNo.

Write the following queries in Relational Algebra. You may use assignment of intermediate results for long queries.

1. Find the name and the SSN of everyone who works more than 100 hours on a project located in Boston.
2. Find the name and SSN of everyone who works for department number 1 and also works on project number 2
3. Find the name and the SSN of everyone who works on at least two projects.
4. Find the name and the SSN of everyone who works on all projects.



Problem 3: Bitmap Index. [27pts]

Consider the following relation Instances:

Instance ID	Storage Type	Storage Size (GB)	Memory Size (GiB)	vCPUs
OW88B1	SSD	475	122	16
3GI579	SSD	950	256	32
HCRNVN	NVMe SSD	475	64	16
INEX1V	NVMe SSD	1900	122	16
B15I2H	HDD	1900	256	64

1. Construct a bitmap index for the attributes ``Storage Size (GB)`` and ``Memory Size (GiB)`` for this table assuming all possible values (except **NULL**) for these two attributes are given in the above table, and the value of these two attributes are allowed to be **NULL**.
2. Assuming now we have N instances, how large space will the above two bitmap indexes occupy?
3. Use the result from the above bitmap indexes and bitwise operations to answer the following query: give the total number of instances that have storage size larger than 800GB and have memory size as 256 GiB.

Typical bitwise operations include `&` (and), `|` (or), `~` (not), `^` (xor), `>> x` (right shift by x bits), `<< x` (left shift by x bits). But we just need to select some of them that can be used here to answer the query.

Assuming the function ``__builtin_popcount(x)`` function in GCC can also be used here to count how many bits are set when the given number is in a binary format.