B561 Assignment 6
Data Generation
Sorting
Indexing
(Draft: Version 1)

Dirk Van Gucht

For this assignment, you will need the material covered in the lectures

- Lecture 15: External Merge Sorting
- Lecture 16: Indexing
- Lecture 17: B^+ trees and Hashing

In addition, you should read the following sections in Chapter 14 and 15 in the textbook *Database Systems The Complete Book* by Garcia-Molina, Ullman, and Widom:

- Section 14.1: Index-structure Basics
- Section 14.2: B-Trees
- Section 14.3: Hash Tables
- Section 14.7: Bitmap Indexes
- Section 15.8.1: Multipass Sort-Based Algorithms

In the file `performingExperiments.sql` supplied with this assignment, we have include several PostgreSQL functions that should be useful for running experiments. Of course, you may wish to write your own functions and/or adopt the functions in this .sql to suite the required experiments for the various problems in this assignment.

The problems that need to be included in the `assignment6.sql` are marked with a blue bullet ●. The problems that need to be included in the `assignment6.pdf` are marked with a red bullet ●. (You should aim to use Latex to construct your .pdf file.) In addition, submit a file `assignment6Code.sql` that contains all the sql code that you developed for this assignment.

Practice problems should not be submitted.

1 Data generation

PostgreSQL functions and clauses In this assignment there will be a need to do simulations with various-size relations. Many of these relations will have randomized data. PostgreSQL provides useful functions and clauses that make such relations:

`random()`	returns a random real number between 0 and 1 using the uniform distribution
`floor(random() * (u - l + 1) + l) :: int`	returns a random integer in the range $[l, u]$ using the uniform distribution
`generate_series(m, n)`	generates the set of integers in the range $[m, n]$
`generate_series(m, n, k)`	generates the set of integers in the range $[m, n]$) separated by step-size k
`order by random()`	randomly orders the tuples that are the result of a query
`limit(n)`	returns the first n tuples from the result of a query
`offset(n)`	returns all but the first n tuples from the result of a query
`limit(m) offset(n)`	returns the first m tuples from all but the first n tuples from the result of a query
`row_number()`	is a *window function* that assigns a sequential integer to each row in a result set
`vacuum`	is a *garbage collection* function to clean and reclaim secondary memory space

For more detail, consult the manual pages

https://www.postgresql.org/docs/13/functions-math.html
https://www.postgresql.org/docs/current/functions-srf.html
https://www.postgresql.org/docs/current/queries-limit.html
https://www.postgresql.org/docs/8.4/functions-window.html
https://www.postgresql.org/docs/9.5/routine-vacuuming.html

In the file `performingExperiments.sql` supplied with this assignment, we have include several PostgreSQL functions that should be useful for running experiments. Of course, you may wish to write your own functions and/or adopt the functions in this .sql to suite the required experiments for the various problems in this assignment.

Generating sets To generate a set, i.e., a unary relation, of n randomly selected integers in the range $[l, u]$, you can use the following function:[1]

```
create or replace function SetOfIntegers(n int, l int, u int)
    returns table (x int) as
$$
    select floor(random() * (u-l+1) + l)::int from generate_series(1,n);
$$ language sql;
```

Example 1 *To generate a unary relation with* 3 *randomly selected integers in the range* 5 *to* 10*, do the following:*

```
select x from SetofIntegers(3,5,10);
```

Of course, running this query multiple times, result in different sets.

[1]Typically the function `SetOfIntegers` returns a bag (multiset) but this is fine for this assignment. In case we want a set, we can always eliminate duplicates.

Generating binary relations The idea behind generating a set can be generalized to that for the generation of a binary relation.[2] To generate a binary relation of n randomly selected pairs of integers (x, y) such $x \in [l_1, u_1]$ and $y \in [l_2, u_2]$, you can use the following function:

```
create or replace function
BinaryRelationOverIntegers(n int, l_1 int, u_1 int, l_2 int, u_2 int)
   returns table (x int, y int) as
$$
   select floor(random() * (u_1-l_1+1) + l_1)::int as x,
          floor(random() * (u_2-l_2+1) + l_2)::int as y
   from   generate_series(1,n);
$$ language sql;
```

Example 2 *To generate a binary relation with* 20 *randomly selected pairs with first components in the range* $[3, 8]$ *and second components in the range* $[2, 11]$*, do the following:*

```
select x, y from BinaryRelationOverIntegers(20,3,8,2,11);
```

Generating functions A relation generated by `BinaryRelationOverIntegers` is in general not a function since it is possible that the relation has pairs (x, y_1) and (x, y_2) with $y_1 \neq y_2$. To create a (partial) *function* $f : [l_1, u_1] \rightarrow [l_2, u_2]$ of n randomly selected function pairs, we can use the following function:

```
create or replace function
FunctionOverIntegers(n int, l_1 int, u_1 int, l_2 int, u_2 int)
   returns table (x int, y int) as
$$
   select x, floor(random() * (u_2-l_2+1) + l_2)::int as y
   from   generate_series(l_1,u_1) x order by random() limit(n);
$$ language sql;
```

Example 3 *To generate a partial function* $[1, 20] \rightarrow [3, 8]$ *of* 15 *randomly selection function pairs, do the following:*[3]

```
select x, y from FunctionOverIntegers(15,1,20,3,8);
```

[2]Clearly, all of this can be generalized to higher-arity relations.

[3]When using this function, it is customary to use n such that $n \in [0, u_1 - l_1 + 1]$.

Generating relations with categorical (non-numeric) data Thus far, the sets, binary relations, and functions have all just involved integer ranges. It is possible to include ranges that have different types including categorical data such as text strings. The technique to accomplish this is to first associate with a categorical range an integer range that associate with each element in the categorical range a unique value of the integer range. The next example illustrates this.

Example 4 *Consider the* `jobSkill` *relation and assume that it contents is*

skill
AI
Databases
Networks
OperatingSystems
Programming

Suppose that we want to generate a `personSkill(pid, skill)` *relation. Let us assume that the* `pid`*'s are integers in the range* $[1, m]$.

There are 5 *skills in the* `jobSkill` *and it is therefore natural to associate with each skill a separate integer (index value) in the range* $[1, 5]$. *This can be done with a query involving the* `row_number()` *window function:*

```
select skill, row_number() over (order by skill) as index
from   Skill;
```

The result is the following relation:

skill	index
AI	1
Databases	2
Networks	3
OperatingSystems	4
Programming	5

Using this technique, we can write a PostgreSQL function that generates a `personSkill` *relation with* n *randomly selected* $(pid, skill)$ *tuples, with* `pid`*'s in the range* $[l, u]$:

```
create or replace function GeneratepersonSkillRelation(n int, l int, u int)
returns table (pid int, skill text) as
$$
with skillNumber(skill, index) as (select skill, row_number() over (order by skill)
                                   from   Skill),
     pS as (select x, y
            from   BinaryRelationOverIntegers(n,l,u,1, (select count(1) from Skill)::int))
select x as pid, skill
from   pS join skillNumber on y = index
group by (x, skill) order by 1,2;
$$ language sql;
```

In this function, the `skillNumber` *view associates with each job skill an integer index in the range* $[1, |\texttt{Skill}|]$. *The* `pS` *view is a randomly generated binary relation with* n *tuples, with* `pid`*'s in the range* $[l, u]$, *and skill numbers in the range* $[1, |\texttt{Skill}|]$. *The* `join` *operation associates the numeric range with the* `Skill` *range. The '*`group by (x, skill) order by 1,2`*' clause eliminates duplicate tuples and orders the result.*

The query

```
select * from GeneratepersonSkillRelation(10,1,15);
```

may make the `personSkill` *relation:*

pid	skill
1	AI
2	Programming
3	Databases
4	Databases
6	Networks
6	OperatingSystems
6	Programming
9	Databases
14	Databases
14	Networks

Problems We now turn to the problems in this section.

1. • Given a discrete probability mass function P, as specified in a relation `P(outcome: int, probability: float)`, over a range of possible outcomes $[u_2, l_2]$, design a PostgreSQL function

   $$\texttt{RelationOverProbabilityFunction}(n, l_1, u_1, l_2, u_2)$$

   that generates a relation of up to n pairs (x, y) such that

   - x is uniformly selected in the range $[l_1, u_1]$; and
   - y is selected in accordance with the probability mass function P in the range $[l_2, u_2]$.

   An example of a possible P as stored in relation `P` is as follows:[4]

   P

outcome	probability
1	0.25
2	0.10
3	0.40
4	0.10
5	0.15

[4]Notice that the sum of the probabilities in the `probability` column in `P` sum to 1.

Note that when P is the uniform probability mass function, then

`RelationOverProbabilityFunction`

and

`BinaryRelationOverIntegers`

are the same binary relation producing functions.

Hint: For insight into this problem, consult the method of *Inverse Transform Sampling* for discrete probability mass functions.

Test your function for the cases listed in the `Assignment-Script-2021-Fall-assignment6.sql` file.

2. **Practice Problem-not graded**.

   Use the technique in Problem 1 and the method for generating categorical data discussed above to write a PostgreSQL function that generates a `personSkill` relation, given a probability mass function over the `Skill` relation.

   Your function should work for any `Skill` relation and any probability distribution defined over it.

Test your function for the cases listed in the `Assignment-Script-2021-Fall-assignment6.sql` file.

2 Sorting

We have learned about *external sorting*. The problems in this section are designed to look into this sorting method as it implemented in PostgreSQL.

3. ● Create successively larger sets of n randomly selected integers in the range $[1, n]$. You can do this using the following function and with the following experiment.[5]

```
create or replace function makeS (n integer)
returns void as
$$
begin
    drop table if exists S;
    create table S (x int);
    insert into S select * from SetOfIntegers(n,1,n);
end;
$$ language plpgsql;
```

   This function generates a bag S of size n, with randomly select integers in the range $[1, n]$. Now consider the following SQL statements:

```
select makeS(10);
explain analyze select x from S;
explain analyze select x from S order by 1;
```

   - The 'select makeS(10)' statement makes a bag S with 10 elements;
   - The 'explain analyze select x from S' statement provides the query plan and execution time in milliseconds (ms) for a simple scan of S;
   - The 'explain analyze select x from S order by 1' statement provides the query plan and execution time in milliseconds (ms) for sorting S.[6]

```
    QUERY PLAN
-----------------------------------------------------------------------------------------------------
 Sort  (cost=179.78..186.16 rows=2550 width=4) (actual time=0.025..0.026 rows=10 loops=1)
   Sort Key: x
   Sort Method: quicksort  Memory: 25kB
   ->  Seq Scan on s  (cost=0.00..35.50 rows=2550 width=4) (actual time=0.004..0.005 rows=10 loops=1)
 Planning Time: 0.069 ms
 Execution Time: 0.034 ms
```

   Now construct the following timing table:[7]

[5] You should make it a habit to use the PostgreSQL vacuum function to perform garbage collection between experiments.

[6] Recall that 1ms is $\frac{1}{1000}$ second.

[7] It is possible that you may not be able to run the experiments for the largest S. If that is the case, just report the results for the smaller sizes.

size n of relation S	avg execution time to scan S (in ms)	avg execution time to sort S (in ms)
10^1		
10^2		
10^3		
10^4		
10^5		
10^6		
10^7		
10^8		

The following tables show the results of experiments for sorting problems, and this for various sizes of the working memory which can be thought of as the buffer size. The experiments were done on a Mac Mini with 2 main-memory modules of 8GB each and 256 SSD disk storage. (The 2 main memory modules permit parallel processing.) All experiments where done with PostgreSQL Version 13.

```
set work_mem = '4MB';
set max_parallel_workers = 0;

 size of relation S | avg execution time to scan S | avg execution time to sort S
--------------------+------------------------------+-----------------------------
                 10 |                        0.009 |                        0.015
                100 |                        0.020 |                        0.044
               1000 |                        0.150 |                        0.384
              10000 |                        1.334 |                        3.668
             100000 |                       13.323 |                       47.916
            1000000 |                      135.008 |                      528.670
           10000000 |                     1387.183 |                     6376.019
          100000000 |                    13985.228 |                    95999.630
```

(a) What are your observations about the query plans for the scanning and sorting of such differently sized bags S?

Solution: Various types of query plans emerged for various setting of the set input sizes. The plans differ in their use of main-memory-only sort (i.e., *quicksort*), *external merge sort*, or *parallel scan and parrallel external merge sort with multiple workers*:

- When the relation S fits in main memory, sorting becomes an in-memory quicksort:

```
                    QUERY PLAN
----------------------------------------------------
 Sort
   Sort Key: x
   Sort Method:  quicksort
   ->  Seq Scan on s
```

- When the relation S does not fit in main memory, sorting is done use the external merge-sort algorithm:

```
                QUERY PLAN
---------------------------------------------
 Sort
   Sort Key: x
   Sort Method:  external merge
   ->  Seq Scan on s
```

- When the relation S does not fit in main memory, parallel external sorting may be done:

```
                               QUERY PLAN
------------------------------------------------------------------------
  Gather Merge
   Workers Planned: 2
   Workers Launched: 2
   ->  Sort
         Sort Key: x
         Sort Method: external merge
          Worker 0:  Sort Method: external merge
          Worker 1:  Sort Method: external merge
         ->   Parallel Seq Scan on s
```

Before showing the results of the experiments, we show the various types of query plans that emerged for various setting of the set input sizes. The plans differ in their use of main-memory-only sort (i.e., *quicksort*), *external merge sort*, or *parallel scan and parrallel external merge sort with multiple workers*:

- When the relation S fits in main memory, sorting becomes an in-memory quicksort:

```
                    QUERY PLAN
----------------------------------------------------
 Sort
   Sort Key: x
   Sort Method:  quicksort
   ->  Seq Scan on s
```

- When the relation S does not fit in main memory, sorting is done use the external merge-sort algorithm:

```
                QUERY PLAN
---------------------------------------------
 Sort
   Sort Key: x
```

```
        Sort Method:  external merge
        ->  Seq Scan on s
```

- When the relation S does not fit in main memory, parallel external sorting may be done:

```
                                   QUERY PLAN
-------------------------------------------------------------------------
 Gather Merge
  Workers Planned: 2
  Workers Launched: 2
   ->  Sort
         Sort Key: x
         Sort Method: external merge
          Worker 0:  Sort Method: external merge
          Worker 1:  Sort Method: external merge
         ->   Parallel Seq Scan on s
```

(b) What do you observe about the execution time to sort `S` as a function of n?

Solution: Recall that the time complexity of external sorting in $O(n \log_B n)$, more precisely $2n\lceil \log_B(n) \rceil$, where n denotes the size of the data and B denotes the size of the main memory buffer.

When we look at the timings for the different main memory sizes, we see that they are consistent with this theoretical complexity analysis. The shape of the curves matching the execution times are of the form $O(n \log_B n)$ and for large buffer size these curves relate to each other in the expected ways. However also note that increasing the buffer size (even dramatically) does not speedup sorting all that much. Again that is predicted by the theory results for sorting. Some more detail:

- For a fixed memory buffer size B, the execution times grow slightly faster than linear, i.e. $O(n)$, which is indeed consistent with the growth rate of an $n \log_B n$ function.
- For a fixed relation size n, we see that the execution times decrease with increasing buffer size. Again, this is consistent with the $n \log_B n$ function as B increases. But note that increasing the buffer size does not have a very dramatic effect. This is because for different buffers size B_1 and B_2, with $B_1 < B_2$, we have
$$n \log_{B_1}(n) = (\log_{B_1}(B_2)) \cdot n \log_{B_2}(n),$$
and for large B_1, $\log_{B_1}(B_2)$ may only grow slowly as a function of B_2.

(c) Does this conform with the formal time complexity of (external) sorting? Explain.

(d) It is possible to set the *working memory* of PostgreSQL using the `set work_mem` command. For example, `set work_mem = '16MB'` sets the working memory to 16MB.[8] The smallest possible working memory in postgreSQL is 64kB and the largest depends on you computer memory size. But you can try for example 1GB.

Repeat question 3a for memory sizes 64kB and 1GB and report your observations.

```
set work_mem = '64kB';
set max_parallel_workers = 0;

size of relation S | avg execution time to scan S | avg execution time to sort S
-------------------+------------------------------+-----------------------------
                10 |                        0.008 |                        0.015
               100 |                        0.020 |                        0.044
              1000 |                        0.149 |                        0.771
             10000 |                        1.337 |                        5.479
            100000 |                       13.371 |                       61.023
           1000000 |                      135.514 |                      858.248
          10000000 |                     1382.106 |                     9819.487
         100000000 |                    14136.311 |                   115182.103
```

```
set work_mem = '4MB';
set max_parallel_workers = 0;

 size of relation S | avg execution time to scan S | avg execution time to sort S
-------------------+------------------------------+-----------------------------
                10 |                        0.009 |                        0.015
               100 |                        0.020 |                        0.044
              1000 |                        0.150 |                        0.384
             10000 |                        1.334 |                        3.668
            100000 |                       13.323 |                       47.916
           1000000 |                      135.008 |                      528.670
          10000000 |                     1387.183 |                     6376.019
         100000000 |                    13985.228 |                    95999.630
```

```
set work_mem = '1GB';
set max_parallel_workers = 0;

size of relation S | avg execution time to scan S | avg execution time to sort S
-------------------+------------------------------+-----------------------------
                10 |                        0.008 |                        0.013
               100 |                        0.018 |                        0.039
              1000 |                        0.133 |                        0.343
             10000 |                        1.342 |                        3.657
            100000 |                       13.600 |                       42.376
           1000000 |                      133.697 |                      494.129
          10000000 |                     1382.028 |                     5512.165
         100000000 |                    14074.251 |                    64347.063
```

We do notice that it takes less time to sort an already sorted relation. It does appear that for sorted data, the complexity is closer to linear $O(n)$, as expected, whereas that for unsorted data, the growth rate in $O(n \log_B n)$.

[8]The default working memory size is 4MB.

(e) Now create a relation `indexedS(x integer)` and create a Btree index on `indexedS` and insert into `indexedS` the sorted relation `S`.[9]

```
create table indexedS (x integer);
create index on indexedS using btree (x);
insert into indexedS select x from S order by 1;
```

Then run the range query

```
select x from indexedS where x between 1 and n;
```

where `n` denotes the size of `S`.

Then construct the following table which contains (a) the average execution time to build the btree index and (2) the average time to run the range query.

```
set work_mem = '4MB';
  size S   | create time for index (in ms) | time for range query (in ms)
-----------+-------------------------------+-----------------------------
        10 |                         0.177 |                        0.015
       100 |                         0.257 |                        0.027
      1000 |                         2.130 |                        0.198
     10000 |                        18.742 |                        1.692
    100000 |                       388.622 |                       17.657
   1000000 |                      3464.479 |                      178.435
  10000000 |                     33583.263 |                     2047.878
 100000000 |                    339423.310 |                    32621.073
```

What are your observations about the query plans and execution times to create `indexedS` and the execution times for sorting the differently sized bags `indexedS`? Compare your answer with those for the above sorting problems.

Solution:

We observe two things:

- The time to create an index on S is expensive. It takes considerably more time to create an index than just sort the relation.
- Since after an B$^+$ tree is created, the leaf level correspond to a sorting in S, retrieving the leaf level to obtain a sort for S is very fast.

Overall, using the strategy to obtain a sorting of S by first indexing and then reading the leaf level of an index is not a good strategy to sort S. However, once the index is created, accessing the sorting of S is clearly beneficial.

[9]For information about *indexes* in PostgreSQL consult the manual page `https://www.postgresql.org/docs/13/indexes.html`.

4. **Practice problem-not graded**.

   Typically, the `makeS` function returns a bag instead of a set. In the problems in this section, you are to conduct experiments to measure the execution times to eliminate duplicates.

   (a) Write a SQL query that uses the `DISTINCT` clause to eliminate duplicates in `S` and report you results in a table such as that in Problem 3a.

   Solution:

```
select distinct from S;

     QUERY PLAN
---------------------
 HashAggregate
   Group Key: x
   ->  Seq Scan on s
(3 rows)

     size  |     execution time
-----------+------------
       100 |      0.05
      1000 |      0.37
     10000 |      3.86
    100000 |     40.11
   1000000 |    662.72
  10000000 |   9720.39
 100000000 | 102204.93
```

   (b) Write a SQL query that uses the `GROUP BY` clause to eliminate duplicates in `S` and report you results in a table such as that in Problem 3a.

   Solution:

```
select x from S group by (x);

     QUERY PLAN
---------------------
 HashAggregate
   Group Key: x
   ->  Seq Scan on s

    size   |     execution time
-----------+------------
```

```
      100 |      0.07
     1000 |      0.50
    10000 |      5.21
   100000 |     57.67
  1000000 |    637.33
 10000000 |   7894.79
100000000 |  97348.35
```

(c) Compare and contrast the results you have obtained in problems 4a and 4b. Again, consider using `explain analyze` to look at query plans.

- We notice that the query plans for the `distinct` and `group by` queries to eliminate duplicates are the same, and therefore the execution times are nearly identical. The query plans reveal that S is put in a hash table on key `x` and this guarantees that duplicates are hashed into the same bucket. The complexity is linear $O(|S|)$.
- The query plan for the `UNION` query reveals that before the hash table is created that S is appended with itself. Consequently a larger relation of size $2|S|$ needs to be hashed. The complexity is linear $O(|S|)$.

3 Indexes and Indexing

Indexes on data (1) permit faster lookup on data items and (2) may speed up query processing on such data. Speedups can be substantial. The purpose of the problems in this section are to explore this. Some other problems in this section are designed to understand the workings of the B^+*-tree* and the *extensible hashing* data structures.

Discussion PostgreSQL permit the creation of a variety of indexes on tables. We will review such `index creation` and examine their impact on data lookup and query processing. For more details, see the PostgreSQL manual:

https://www.postgresql.org/docs/13/indexes.html

Example 5 *The following SQL statements create indexes on columns or combinations of columns of the* `personSkill` *relation.*[10] *Notice that there are*

$$2^{arity(\texttt{personSkill})} - 1 = 2^2 - 1 = 3$$

such possible indexes.

```
create index pid_index on personSkill (pid);                -- index on pid attribute
create index skill_index on personSkill (skill);            -- index on skill attribute
create index pid_skill_index on personSkill (pid,skill);    -- index (pid, skill)
```

Example 6 *It is possible to declare the type of index:* `btree` *or* `hash`*. When no index type is specified, the default is* `btree`*. If instead of a Btree, a* hash index *is desired, then it is necessary to specify a* `using hash` *qualifier:*

```
create index pid_hash on personSkill using hash (pid);  -- hash index on pid attribute
```

Example 7 *It is possible to create an index on a relation based on a* scalar expression *or a* function *defined over the attributes of that relation. Consider the following (immutable) function which computes the number of skills of a person:*

```
create or replace function numberOfSkills(p integer) returns integer as
$$
    select count(1)::int
    from   personSkill
    where  pid = p;
$$ language SQL immutable;
```

[10] Incidentally, when a primary key is declared when a relation is created, PostgreSQL will create a btree index on this key for the relation.

Then the following is an index defined on the `numberOfSkills` *values of persons:*

```
create index numberOfSkills_index on personSkill (numberOfSkills(pid));
```

Such an index is useful for queries that use this function such as

```
select pid, skill from personSkill where numberOfSkills(pid) > 2;
```

We now turn to the problems in this section.

5. ● Consider a relation `Student(sid text, sname text, major, year)`. A tuple (s, n, m, y) is in `Student` when s is the sid of a student and n, m, and y are that student's name, major, and birth year. Further, consider a relation `Enroll(sid text, cno text, grade text)`. A triple (s, c, g) is in `Enroll` when the student with sid s was enrolled in the course with cname c and obtained a letter grade g in that course.

   We are interested in answering queries of the form

```
select sid, sname, major, byear
from   Student
where  sid in (select sid
               from   Enroll sid
               where  cno = c [and|or|and not] grade = g);
```

   Here `c` denotes a course name and `g` denotes a letter grade.

   Read Section 14.1.7 'Indirection in Secondary Indexes' in your textbook *Database Systems The Complete Book* by Garcia-Molina, Ullman, and Widom. Of particular interest are (a) the concept of *buckets* (Figure 14.7) which are sets of tids and (b) the technique of performing set operations (like intersections) on relevant buckets (Figure 14.8) to answer queries of the form as shown above.

   The goal of this problem is to use object-relational SQL to simulate these concepts. To make things more concrete, consider the following `Student` and `Enroll` relations:

```
Student:
 sid  | sname  |  major  | byear
------+--------+---------+-------
 s100 | Eric   | CS      |  1988
 s101 | Nick   | Math    |  1991
 s102 | Chris  | Biology |  1977
 s103 | Dinska | CS      |  1978
 s104 | Zanna  | Math    |  2001
 s105 | Vince  | CS      |  2001

Enroll:
 sid  | cno  | grade
------+------+-------
 s100 | c200 | A
 s100 | c201 | B
 s100 | c202 | A
 s101 | c200 | B
 s101 | c201 | A
 s101 | c202 | A
 s101 | c301 | C
 s101 | c302 | A
 s102 | c200 | B
 s102 | c202 | A
 s102 | c301 | B
 s102 | c302 | A
 s103 | c201 | A
 s104 | c201 | D
```

Now consider associating a tuple id (tid) with each tuple in `Enroll`:

```
 tid | sid  | cno  | grade
-----+------+------+-------
   1 | s100 | c200 | A
   2 | s100 | c201 | B
   3 | s100 | c202 | A
   4 | s101 | c200 | B
   5 | s101 | c201 | A
   6 | s101 | c202 | A
   7 | s101 | c301 | C
   8 | s101 | c302 | A
   9 | s102 | c200 | B
  10 | s102 | c202 | A
  11 | s102 | c301 | B
  12 | s102 | c302 | A
  13 | s103 | c201 | A
  14 | s104 | c201 | D
```

Use object-relational SQL to construct two secondary indexes `indexOnCno` and `indexOnGrade` on the `Enroll` relation. These indexes should be stored in two separate relations which you can conveniently call `indexOnCno` and `indexOnGrade`, respectively. two object-relational views in a manner that simulates the situation illustrated in Figure 14.8. In particular, do **not** use the '`create index`' mechanism of SQL to construct these indexes.

Then, using the `indexOnCno` and `indexOnGrade` views and the technique of *intersecting buckets*, write a function `FindStudents(booleanOperation text, cno text, grade text)` that can be used to answer queries of the form as shown above. (Here the booleanOperation is a string which can be 'and', 'or', or 'and not'.)
For example, the query

```
select * from FindStudents('and', 'c202', 'A');
```

should return the same result as that of the query

```
select sid, sname, major, byear
from   Student
where  sid in (select sid
               from   Enroll sid
               where  cno = 'c202' and grade = 'A');
```

Test your solution for the following cases on the `Student` and `Enroll` relation given for this problem:

(a) `select * from FindStudents('and', 'c202', 'A');`

(b) `select * from FindStudents('or', 'c202', 'A');`

(c) `select * from FindStudents('and not', 'c202', 'A');`

6. **Practice problem–not graded**. Read Section 14.7 'Bitmap Indexes' in your textbook *Database Systems The Complete Book* by Garcia-Molina, Ullman, and Widom. In particular, look at Example 14.39 for an example of a bitmap index for a secondary index.

Next, revisit Problem 5. There, we considered two secondary indexes `indexOnCno` and `indexOnGrade`. We will now consider the corresponding bitmap indexes `bitmapIndexOnCno` and `bitmapIndexOnGrade`:

```
bitmapIndexOnCno
 cno  |     bit-vector
------+----------------
 c200 | 10010000100000
 c201 | 01001000000011
 c202 | 00100100010000
 c301 | 00000010001000
 c302 | 00000001000100
```

and

```
bitmapIndexOnGrade
 grade |     bit-vector
-------+----------------
 A     | 10101101010110
 B     | 01010000101000
 C     | 00000010000000
 D     | 00000000000001
```

Use object-relational SQL to construct two secondary indexes `bitmapIndexOnCno` and `bitmapIndexOnGrade` as two object-relational relations in a manner that simulates the situation just illustrated above.

Then, using the `bitmapIndexOnCno` and `bitmapIndexOnGrade` relations and the technique of forming the bitmap-`AND`, bitmap-`OR`, and bitmap-`AND NOT` of two bit-vectors, write a function `FindStudents(booleanOperation text, cno text, grade text)` that can be used to answer queries of the form as shown in Problem 5.
For example, the query

```
select * from FindStudents('and', 'c202', 'A');
```

should return the same result as that of the query

```
select sid, sname, major, byear
from   Student
where  sid in (select sid
               from   Enroll sid
               where  cno = 'c202' and grade = 'A');
```

Test your solution for the following cases on the `Student` and `Enroll` relation given for this problem:

(a) `select * from FindStudents('and', 'c202', 'A');`

(b) `select * from FindStudents('or', 'c202', 'A');`

(c) `select * from FindStudents('and not', 'c202', 'A');`

7. ● Consider the following parameters:

block size	=	4096 bytes
block-address size	=	9 bytes
block access time (I/O operation)	=	10 ms (micro seconds)
record size	=	150 bytes
record primary key size	=	8 bytes

Assume that there is a B$^+$-tree, adhering to these parameters, that indexes 1 billion (10^9) records on their primary key values.

Provide answers to the following problems and show the intermediate computations leading to your answers.

(a) Specify (in ms) the minimum time to determine if there is a record with key k in the B$^+$-tree. (In this problem, you can not assume that a key value that appears in a non-leaf node has a corresponding record with that key value.)

Solution: We first need to determine the order n of the B$^+$-tree. This can be done by finding the largest n such that

$$9(n+1) + 8n \leq 4096$$

I.e., $n \leq \lfloor \frac{4096-9}{17} \rfloor$. Thus $n = 240$.

Note that there will be 10^9 keys at the leaf level of the B$^+$ tree. The keys of the 10^9 records can be stored in $\frac{10^9}{240} = 4166667$ leaf nodes of the B^+ tree.

Note that the minimum time is determined by the largest possible fanout of the nodes in the B$^+$-tree. This fanout is maximally $n+1 = 240+1 = 241$.

The minimum time will therefore be $(\lceil log_{241}(4166667) \rceil) * 10ms = 3 * 10\,ms = 30\,ms$.

(b) Specify (in ms) the maximum time to insert a record with key k in the B$^+$-tree assuming that this record was not already in the data file.

Solution: The maximum time results when we have the minimum branching factor, i.e. $\lceil \frac{240}{2} \rceil + 1 = 120 + 1 = 121$ at non-root nodes.

Note that this time, there will be $\frac{10^9}{120}$ leaf nodes of the B$^+$-tree that store the keys of the 10^9 keys of the records.

Since the minimum branching factor at the root is 2, we must determine the height of a tree that has half of the nodes, i.e., $\lceil \frac{\frac{10^9}{120}}{2} \rceil =$ 4166667 nodes.

The height of such a the tree will be $\lceil log_{121}(4166667) \rceil = \lceil 3.178 \cdots \rceil =$ 4.

We then also need one more block access to insert the record and another to write that block back to the data file, so the maximum time is $(4+2+1)10\,ms = 70\,ms$. The 1 in this sum is present because we also need to retrieve the root block.

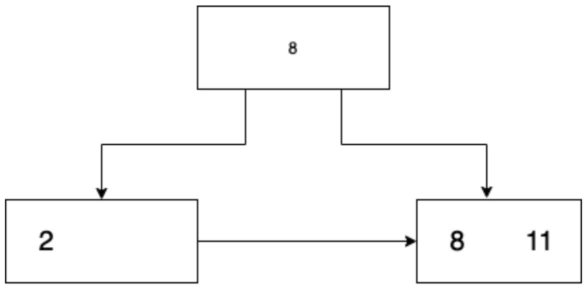
(c) How large must main memory be to hold the first two levels of the B^+-tree? How about the first three levels?

Solution:

We must store the root and the next level which may have many as 241 subtrees (i.e., when we have maximum branching at the root). This give a total of $1 + 241$ blocks which is about 1MB of space in main memory to store the 2 top levels of the B^+ tree

If we need to store the first 3 level, we need have a need for $1+241+241^2 = 58323$ blocks which is about 0.25GB.

8. ● Consider the following B$^+$-tree of order 2 that holds records with keys 2, 8, and 11.[11]



(a) Show the contents of your B^+-tree after inserting records with keys 4, 10, 12, 1, 7, and 5 in that order.

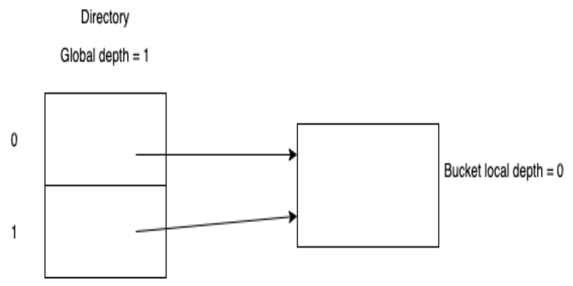
Strategy for splitting leaf nodes: when a leaf node n needs to be split into two nodes n_1 and n_2 (where n_1 will point to n_2), then use the rule that an even number of keys in n is moved into n_1 and an odd number of keys is moved into n_2. So if n becomes conceptually $\boxed{k_1|k_2|k_3}$ then n_1 should be $\boxed{k_1|k_2}$ and n_2 should be $\boxed{k_3|\quad}$ and $n_1 \rightarrow n_2$.

(b) Starting from your answer in question 8a, show the contents of your B^+-tree after deleting records with keys 12, 2, and 11, in that order.

(c) Starting from your answer in question 8b, show the contents of your B^+-tree after deleting records with keys 5, 1, and 4, in that order.

[11] Recall that (a) an internal node of a B$^+$-tree of order 2 can have either 1 or 2 keys values, and 2 or 3 sub-trees, and (b) a leaf node can have either 1 or 2 key values.

9. ● Consider an extensible hashing data structure wherein (1) the initial global depth is set at 1 and (2) all directory pointers point to the same **empty** bucket which has local depth 0. So the hashing structure looks like this: Assume that a bucket can hold at most two records.



(a) Show the state of the hash data structure after each of the following insert sequences:[12]

i. records with keys 6 and 7.
ii. records with keys 1 and 2.
iii. records with keys 9 and 4.
iv. records with keys 8 and 3.

(b) Starting from the answer you obtained for Question 9a, show the state of the hash data structure after each of the following delete sequences:

i. records with keys 9 and 6.
ii. records with keys 4 and 1.
iii. records with keys 2 and 8.

As in the text book, the bit strings are interpreted starting from their left-most bit and continuing to the right subsequent bits.

[12]You should interpret the key values as bit strings of length 4. So for example, key value 7 is represented as the bit string 0111 and key value 2 is represented as the bit string 0010.

The goal in the next problems is to study the behavior of indexes for various different sized instances[13] of the `Person`, `worksFor`, and `Knows` relations and for various queries:

10. ● Create an appropriate index on the `worksFor` relation that speedups the lookup query

```
select pid from worksFor where cname = c;
```

Here c is a company name.

Illustrate this speedup by finding the execution times for this query for various sizes of the `worksFor` relation.

Solution

We use a hash secondary index on name and compare scanning and hash index search for different sizes of the `worksFor` relation.

```
create index index_on_cname on worksFor using btree (cname);

We assume 10000 different companies and 20 different salaries.

worksFor      |sequential search| index search
1000000       |   80.300        | 0.029
10000000      |  923.536        | 0.295
```

We observe significant improvement with hash index search as expected. (Note that for larger data, the query needs to output more pids which explain the different for the hash index search.).

[13](Three different sizes, small, medium, large suffice for your experiment.

11. ● Create an appropriate index on the `worksFor` relation that speedups the range query

```
select pid, cname
from   worksFor
where  s1 <= salary and salary <= s2;
```

Here s1 and s2 are two salaries with s1 < s2.

Illustrate this speedup by finding the execution times for this query for various sizes of the `worksFor` relation.

Solution

We use a B^+-tree and compare scanning and range index search for different size of the worksFor relation and for different ranges.

```
create index index_on_salary on worksFor using btree (salary);

We assume 10000 different companies and 20 different salaries.

worksFor     |sequential search| index search (range 5000,5000)
1000000      |   80.979        | 0.067
10000000     |  947.967        | 0.335

worksFor     |sequential search| index search (range 5000,6000)
1000000      |   80.979        | 20.221
10000000     |  947.967        | 207.910

worksFor     |sequential search| index search (full range)
1000000      |   80.979        | 170.339
10000000     |  947.967        | 1656.778 ms
```

We observe significant improvement with B^+-tree index search for small ranges, moderate improvement for medium ranges, and a slowdown for full range. We observe that a sequential scan is more efficient than an index scan over the full range. The reason is that, in a full range index scan, two phases happen, a scan through the index to determine the tids of records that qualify, and subsequently a 'lookup' scan to access to records in the data file.

12. ● Create indexes on the `worksFor` relation that speedup the multiple conditions query

```
select pid
from   worksFor
where  salary = s and cname = c;
```

Here s is a salary and c is a company name.

Illustrate this speedup by finding the execution times for this query for various sizes of the `worksFor` relation.

Solution

- Approach 1: with btrees

  We use two btrees for the secondary indexes on salaries and companies, respectively.

  ```
  create index index_on_cname on worksFor using btree (cname);
  create index index_on_salary on worksFor using btree (salary);
  ```

  We compare sequential scanning and index search for different sizes of the worksFor relation. We assume 10000 different companies and 20 different salaries

  ```
  worksFor     |sequential search| index search
  1000000      |   81.596        |   0.033
  10000000     |  978.551        |  72.056
  ```

  Notice query plan. The index search is on a salary which, for our data, means that there are lots of employees with that salary. Only after these persons are found, there is a filter condition on cname.

  ```
   Index Scan using index_on_salary on worksfor
     Index Cond: (salary = 10000)
     Filter: (cname = 5000)
  ```

- Approach 2: with hash indexes

  We use two hash indexes for the secondary indexes on salaries and companies, respectively.

  ```
  create index index_on_cname on worksFor using hash (cname);
  create index index_on_salary on worksFor using hash (salary);
  ```

  We compare sequential scanning and index search for different sizes of the worksFor relation. We assume 10000 different companies and 20 different salaries

  ```
  worksFor     |sequential search| index search
  1000000      |   81.596        | 0.033
  10000000     |  966.913 ms     | 0.850
  ```

Consider the query plan. The index search, which is a bitmap index scan, is on company cname. For a particular company, there are few employees who work for that company. (Note that the relevant tids are collected as a bitmap.) The filter condition checks for salary is therefore very fast.

```
 Bitmap Heap Scan on worksfor
   Recheck Cond: (cname = 5000)
   Filter: (salary = 10000)
   ->  Bitmap Index Scan on index_on_cname
         Index Cond: (cname = 5000)
```

This different choice of the chosen index (for btree, the salary index; for hash index, the cname index) explains the difference in execution time between the btree and hash data structure.

It it is also evident that in both approaches, the secondary indexes are not implemented using the buckets of tids.

Still, in both case, we observe significant improvement relative to sequential scanning. In our experiments, hash indexing performs better.

- In the following experiment, we try to balance out the number of different companies and different salaries. We set both to 400 different values each. We get a new insight: now the buckets intersection strategy kicks (using bitmaps) in as the following query plan demonstrates. We thus get symmetric handling of the two secondary indexes.

```
 Bitmap Heap Scan on worksfor
   Recheck Cond: ((salary = 100000) AND (cname = 50))
   ->  BitmapAnd
         ->  Bitmap Index Scan on index_on_salary
               Index Cond: (salary = 100000)
         ->  Bitmap Index Scan on index_on_cname
               Index Cond: (cname = 50)

worksFor     |sequential search| index search (with btrees)
1000000      |   83.393        | 0.098
10000000     |  961.040        | 0.919

worksFor     |sequential search| index search (with hash)
1000000      |   83.393        | 0.232
10000000     |  961.040        | 4.464
```

We observe significant speedup for either btrees and hash indexes. The btrees appear superior for this data.

13. ● Create indexes on the appropriate relations that speedup the semi-join [anti semi-join] query

```
select pid, pname
from   Person
where  pid [not] in (select pid from worksFor where cname = c);
```

Here c is a company name.

Illustrate this speedup by finding the execution times for this query for various sizes of the Person and worksFor relations.

Solution

We create a secondary index for worksFor on salaries. We create a primary index for Person on pids.

We compare the performance of running the queries without indexes (an hash join algorithm is used) and with indexes. The performances are all very similar.

```
Person   | Worksfor | Sequential (hash join) | With indexes (hash index)
2000000  |1000000   | 471.765                | 406.80   (for IN query)
2000000  |1000000   | 474.765                | 434.160  (for NOT IN query)
```

We compare the query plans:

- IN query

  The algorithm without index is a simple hash join algorithm with hash table creation on $\sigma_{cname=50}(worksfor)$.

```
 Hash Join
   Hash Cond: (person.pid = worksfor.pid)
   ->  Seq Scan on person
   ->  Hash
         ->  HashAggregate
               Group Key: worksfor.pid
                   ->  Parallel Seq Scan on worksfor
                         Filter: (cname = 50)
```

  For join algorithm with index (note however how this query plan ignores the index for person on pid).
  This query plan is almost the same as that one above, but is uses bitmaps following consulting the index for worksFor on cname before building a hash table.

```
   ->  Hash Semi Join
         Hash Cond: (person.pid = worksfor.pid)
         ->  Seq Scan on person
         ->  Hash
               ->  Bitmap Heap Scan on worksfor
                     Recheck Cond: (cname = 50)
                       ->  Bitmap Index Scan on index_on_cname
                             Index Cond: (cname = 50)
```

- `NOT IN` query

  For hash-lookup algorithm with hash-table creation on $\sigma_{cname=50}(worksFor)$. This is essentially an anti-join algorithm. It is very similar to the semi-join algorithm.

  ```
  Seq Scan on person
     Filter: (NOT (hashed SubPlan 1))
     SubPlan 1
       ->  Parallel Seq Scan on worksfor
                 Filter: (cname = 50)
  ```

  For join algorithm with index (note however how this query plan ignores the index for person on pid).

  This query plan is almost the same as that one above. Also not how similar it is to the query plan for the `IN` query.

  ```
  Hash Join
     Hash Cond: (person.pid = worksfor.pid)
     ->  Seq Scan on person
     ->  Hash
           ->  HashAggregate
                 Group Key: worksfor.pid
                 ->  Bitmap Heap Scan on worksfor
                       Recheck Cond: (cname = 50)
                          ->  Bitmap Index Scan on index_on_cname
                              Index Cond: (cname = 50)
  ```

Looking at the 4 query plan, we note that they nearly identical, so we can expect their execution time to be very similar. It would appear that bitmap creation speedup the queries marginally.

14. **Practice problem–not graded.**

Create indexes that speedup the path-discovery query

```
select distinct k1.pid1, k3.pid2
from   knows k1, knows k2, knows k3
where  k1.pid2 = k2.pid1 and k2.pid2 = k3.pid1;
```

Illustrate this speedup by finding the execution times for this query for various sizes of the `Knows` relation.

Solution

```
create index index_on_pid1 on knows using btree (pid1);
create index index_on_pid2 on knows using btree (pid2);

Knows has 10000 pairs 1000 pid1, 10 pid2
Knows has 100000 pairs 10000pid2, 10 pid2

Knows    | sequential search|  index search
10000    |     626.538      |  514.615
100000   |    4068.140      | 2720.942
```

We observe marginal improvement.