October 25, 2018
Arnab Chakrabarti, Rihan Hai

Implementation of Databases (WS 18/19)

Exercise 2

Due until November 6, 2018, 10am.
Please submit your solution *in a single PDF file* before the deadline to the L^{2}P system!
Please submit solutions in groups of three students.

Exercise 2.1 (Relational Calculus) (8 pts)

Suppose you have the following database schema:

employee(emp_id, name, salary)
flights(flight_no, from, to, distance, depart_time, arrival_time)
aircraft(aircraft_id, manufacturer, model, range)
certified(emp_id, aircraft_id)

Represent the following queries as **tuple relational calculus** (TRC), **domain relational calculus** (DRC).

[Note: The certified relation indicates which employee(s) is/are certified to fly which aircraft.]

1. Find the names of employees who are certified to fly aircraft manufactured by Boeing.
2. Find the aircraft_ids of all aircrafts that can be used on non-stop flights (i.e. where the aircraft.range > flights.distance) from Vancouver to Tokyo.
3. Find the names of pilots who can operate planes with a range greater than 3000 miles but are not certified on any aircraft manufactured by Boeing.
4. Find the employee id's of the employees who make the highest salary.

Exercise 2.2 (Sorting)

Suppose you have a file of 20,000 pages and five buffer pages and you are sorting it using general (external) merge-sort. Please answer the following questions: **(4 pts)**

1. How many runs will you produce? Remark: When a file is sorted, in intermediate steps subfiles are created. Each sorted subfile is called a run. (1 pt.)
2. How many passes will it take to sort the file completely? (1 pt.)
3. How many buffer pages do you need at least to sort the file in two passes? (2 pt.)

Exercise 2.3 (Indexing) **(8 pts)**

On the relation Cities(Name, Province, Population, KilometresFromAachen) we generate the following 2 queries:

- Q1:

```
SELECT Name, Province
FROM Cities
WHERE Population > 100000
```

- Q2:

```
SELECT Province, count(city), sum(Population)
FROM Cities
group by Province
```

Now answer the following questions:

1. Briefly explain how a B+-tree index on Population could be used during processing of Q1.
2. Briefly explain why using a clustered index on Province would be more efficient than either a hash-table or B+-tree index during processing of Q2.

Exercise 2.4 (Short Questions) **(10 pts)**

Answer briefly the following questions:

1. Give two examples of SQL constructs/semantics not expressible in relational algebra (RA). (2 pt.)
2. Explain the difference between DRC and TRC. (2 pt.)
3. What does "*relational completeness*" mean (in your own words, please)? Show that SQL is relationally complete by enumerating SQL constructs corresponding to selection, projection, cartesian product, union, and difference. (3 pt.)
4. What is an unsafe query? Considering the schema given in exercise 1.1, give an example and explain why it is important to disallow such queries. (3 pt.)