ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Distributed Computing

FS 2019 Prof. R. Wattenhofer

Computer Engineering II

Solution to Exercise Sheet Chapter 7

1 Quiz Questions

a) Both {`ID,TR-ID`} and {`ID,title,TR-ID`} are superkeys, because they uniquely identify any row within the table. As there are two rows with the same `ID` and two rows with the same `TR-ID`, it is not sufficient to use only one of these two columns to identify a row. Hence, {`ID,TR-ID`} is a candidate key. Thus, {`ID,title,TR-ID`} is not a candidate key, because the `title` column can be omitted.

b) Query 6. results in:

```
ERROR 1111 (HY000):  Invalid use of group function.
```

SQL's WHERE clause does not work with aggregate functions like SUM, AVG, MAX, COUNT and so on. Instead, the HAVING keyword was introduced to SQL in order to quantitatively compare aggregateed values. A correct query would look like this:

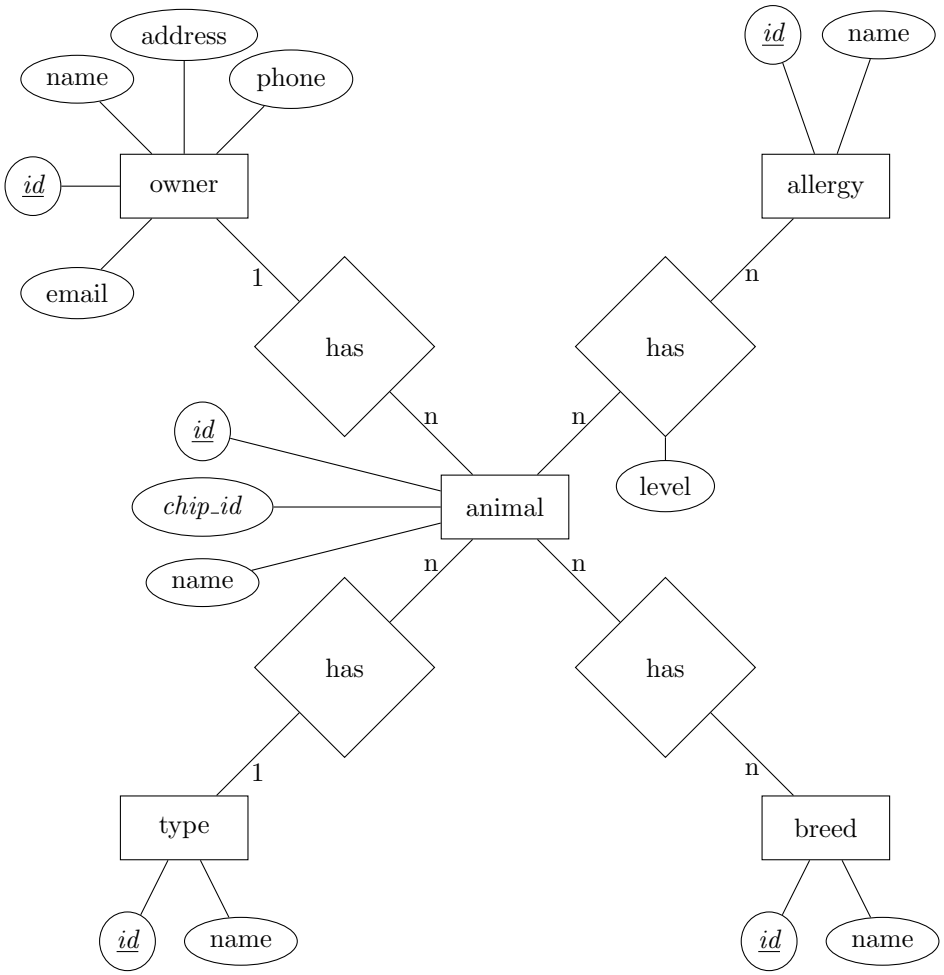
```
SELECT year, COUNT(*) FROM Articles GROUP BY year HAVING COUNT(*) > 10;
```

c) 3 rows

d) The result of a right outer join includes all the rows from Table R, with the corresponding detail of Table L for those rows for which the value of the two tables' common attribute match exactly.

2 Database Design

a) The following Entity Relationship Diagram describes the animals database. Owners and animals are in a 1-to-n relation. Each owner may own multiple animals, but every animal can have exactly one registered owner in the database. Animals and animal types are in a n-to-1 relation. Any animal cannot be both a cat and a dog, but the animal type table may very well contain multiple cats or dogs. Animals and breeds are in a n-to-n relation. Any animal can be a mixed breed and there may be multiple animals of the same breed in the database. Animals and allergies are in a n-to-n relation. Any animal may have multiple allergies and any allergy may afflict more than one animal in the database. For every animal allergy, we reserve a level field that denotes how strongly allergic the animal is to the allergy in question. Notice that we underline primary key attributes and we use an italic font to label unique attributes. The id field is chosen as the primary key for the animal table because even though the chip_id is a unique value, we would like to allow for the possibility that the unique chip_id is changed, for example, when the animal chip breaks or if it's updated due to a change in chip standard.



b)

```
DROP TABLE IF EXISTS Owner;
CREATE TABLE Owner (
id INT PRIMARY KEY,
name CHARACTER VARYING(255),
email CHARACTER VARYING(255),
phone CHARACTER VARYING(15),
address CHARACTER VARYING(255)
);

DROP TABLE IF EXISTS Type;
CREATE TABLE Type (
id INT PRIMARY KEY,
animal_type CHARACTER VARYING(50)
);


DROP TABLE IF EXISTS Animal;
CREATE TABLE Animal (
id INT PRIMARY KEY,
chip_id CHARACTER VARYING(50) UNIQUE,
name CHARACTER VARYING(50),
```

```
owner_id INT,
type_id INT,
CONSTRAINT fk_animal_owner FOREIGN KEY (owner_id) REFERENCES Owner(id),
CONSTRAINT fk_animal_type FOREIGN KEY (type_id) REFERENCES Type(id)
);

DROP TABLE IF EXISTS Allergy;
CREATE TABLE Allergy (
id INT PRIMARY KEY,
allergy CHARACTER VARYING(255)
);

DROP TABLE IF EXISTS AllergicAnimals;
CREATE TABLE AllergicAnimals (
id INT PRIMARY KEY,
allergy_degree INT,
animal_id INT,
allergy_id INT,
CONSTRAINT fk_allergic_animal_id FOREIGN KEY (animal_id) REFERENCES Animal(id),
CONSTRAINT fk_allergy_id FOREIGN KEY (allergy_id) REFERENCES Allergy(id)
);

DROP TABLE IF EXISTS Breed;
CREATE TABLE Breed (
id INT PRIMARY KEY,
animal_breed CHARACTER VARYING(50)
);

DROP TABLE IF EXISTS AnimalBreed;
CREATE TABLE AnimalBreed (
id INT PRIMARY KEY,
animal_id INT,
breed_id INT,
CONSTRAINT fk_animal_breed_animal_id FOREIGN KEY (animal_id) REFERENCES Animal(id),
CONSTRAINT fk_animal_breed_breed_id FOREIGN KEY (breed_id) REFERENCES Breed(id)
);
```

3 Database Queries

1. `SELECT id,title FROM movie LIMIT 5;`
2. `SELECT * FROM movie ORDER BY title DESC LIMIT 2;`
3. `SELECT COUNT(*) FROM movie WHERE year > 2000;`
4. `SELECT title,tomatometer FROM movie WHERE title = 'The Matrix';`
5.

```
SELECT COUNT(*) FROM movie
  WHERE tomatometer > ALL (
    SELECT tomatometer FROM movie
      WHERE title = 'The Matrix'
    );
```

6.
```
SELECT MAX(tomatometer),MIN(tomatometer),AVG(tomatometer) FROM movie;
```

7.
```
SELECT year, AVG(tomatometer) AS avg FROM movie
GROUP BY year
ORDER BY avg DESC LIMIT 5;
```

8.
```
SELECT title FROM movie
  WHERE title LIKE 'X%'
  ORDER BY title DESC;
```

9.
```
SELECT COUNT(*) FROM movie
  WHERE title LIKE '%fight%';
```

   or case insensitive:

```
SELECT COUNT(*) FROM movie
  WHERE title COLLATE UTF8_GENERAL_CI LIKE '%fight%';
```

10.
```
SELECT title FROM movie
WHERE title LIKE '%X%M%' OR title LIKE '%M%X%';
```

4 B+ Trees

1. Since both the root and the internal nodes of the tree have at least 2 children, the tree has at least 4 leaves. Each of these leaves contain at least 1 key, so the minimum possible number of keys is 4.
   As for the maximum, both the root and the internal nodes have at most 4 children, so the maximal number of leaves is 16. Each of these leaves can store up to 3 keys, summing up to a maximum of 48 leaves.

2. Since each node has the maximum number of children, the splitting operation will propagate up to the root, increasing the depth of the tree by 1. The new tree will have 2 internal nodes at depth one, and 5 internal nodes at depth two, so 7 altogether.

3. All nodes have the minimum number of children, so the merging operation will again propagate up, and merge the two children of the root. With that, the original root is discarded, and the depth of the tree will decrease by one. The resulting tree will have no internal nodes at all.

5 More Database Queries

1.

```
SELECT person.name,cast_info.role_id,person.gender FROM cast_info
    JOIN person ON person.id = cast_info.person_id
    JOIN movie ON movie.id = cast_info.movie_id
JOIN role_type ON role_type.id = cast_info.role_id
  WHERE role_type.role = 'actress'
    AND movie.title = 'The Matrix';
```

2.

```
SELECT COUNT(DISTINCT person.id) FROM cast_info
    JOIN role_type ON role_type.id = cast_info.role_id
    JOIN person ON person.id = cast_info.person_id
  WHERE role_type.role = 'director'
  AND person.gender = 'f';
```

3.

```
SELECT DISTINCT person.name FROM cast_info
    JOIN person ON person.id = cast_info.person_id
    JOIN movie ON movie.id = cast_info.movie_id
  WHERE (cast_info.role_id = 2 or cast_info.role_id = 1)
  AND EXISTS (
    SELECT DISTINCT ci.person_id FROM cast_info AS ci
      WHERE ci.role_id = 8
      AND cast_info.person_id = ci.person_id
      GROUP BY ci.person_id
      HAVING COUNT(ci.person_id) > 20
  );
```

Alternative solution:

```
SELECT DISTINCT person.name FROM person
  JOIN cast_info ON person.id = cast_info.person_id
  JOIN role_type ON cast_info.role_id=role_type.id
WHERE role_type.role IN ('actor','actress')
AND 20 < (
  SELECT COUNT(*) FROM cast_info AS ci
    JOIN role_type AS rt ON ci.role_id=rt.id
    WHERE ci.person_id = person.id
    AND rt.role='director'
);
```

4.

```
SELECT movie.title,COUNT(*) AS cnt FROM movie_keyword
    JOIN movie ON movie_keyword.movie_id = movie.id
  GROUP BY movie_keyword.movie_id
  ORDER BY cnt DESC;
```

5.

```
SELECT AVG(cnt),MAX(cnt),MIN(cnt) FROM (
  SELECT movie.title,COUNT(*) AS cnt FROM movie_keyword
      JOIN movie ON movie_keyword.movie_id = movie.id
    GROUP BY movie_keyword.movie_id
  ) AS countaverages;
```

6.

```
SELECT
    person.name,
    AVG(movie.tomatometer) AS average,
    COUNT(ci.person_id) AS cnt,
    MAX(movie.year) AS maxyear
  FROM cast_info AS ci
    JOIN movie ON movie.id = ci.movie_id
    JOIN person ON person.id = ci.person_id
  WHERE ci.role_id = 1
  GROUP BY ci.person_id
  HAVING average > 85
  AND cnt > 30
  AND maxyear > 2000
  ORDER BY
    maxyear DESC,
    average DESC;
```

7.

```
SELECT person.name FROM person
    JOIN cast_info ON person.id = cast_info.person_id
    JOIN movie ON cast_info.movie_id = movie.id
  WHERE cast_info.role_id = 8
  AND movie.tomatometer > 90
  GROUP BY person.id
  HAVING COUNT(*) > 10;
```