Distributed Data Management
Dr. Eike Schallehn / Ingolf Geist

Exercise 7 - Distributed Query Processing I

based on [1]

The following relation schema is given:

```
EMPLOYEE(ENR, ENAME, JOB, SALARY)
PROJECT(PNR, PNAME, BUDGET)
ASSIGNMENT(ENR, PNR, DURATION)
```

1. Normalization

What is the disjunctive and conjunctive normal form for the `WHERE`-condition of the following query:

```
SELECT * FROM EMPLOYEE
WHERE (ENAME LIKE 'M%' AND JOB='Administrator')
OR ((ENR>550 OR JOB='SW-Developer') AND SALARY<80.000)
```

2. Simplification

Simplify the selection condition of the following query:

```
SELECT * FROM EMPLOYEE
WHERE ENR>456 AND
NOT (JOB='Administrator' OR SALARY<50.000) AND
JOB<>'Administrator' AND SALARY<50.000
```

3. Operator Tree and Fragment Expression

Transform the following query to relational algebra and perform algebraic optimizations:

```
SELECT ENAME, PNAME
FROM EMPLOYEE P, PROJECT PT, ASSIGNMENT PM
WHERE DURATION>10 AND P.ENR=PM.ENR AND
JOB='SW-Developer' AND PT.PNR=PM.PNR
```

4. Data Localization)

The following horizontal fragmentation for PROJECT is given:

$$\text{PROJECT}_1 = \sigma_{\text{BUDGET}<100.000}(\text{PROJECT})$$
$$\text{PROJECT}_2 = \sigma_{100.000\leq\text{BUDGET}\leq800.000}(\text{PROJECT})$$
$$\text{PROJECT}_3 = \sigma_{\text{BUDGET}>800.000}(\text{PROJECT})$$

ASSIGNMENT has an according derived fragmentation:

$$\text{ASSIGNMENT}_1 = \text{ASSIGNMENT} \ltimes \text{PROJECT}_1$$
$$\text{ASSIGNMENT}_2 = \text{ASSIGNMENT} \ltimes \text{PROJECT}_2$$
$$\text{ASSIGNMENT}_3 = \text{ASSIGNMENT} \ltimes \text{PROJECT}_3$$

Transform the following query to relational algebra and create the initial fragment expressions by inserting the reconstruction expression. Perform algebraic optimization to improve the fragment expression.

```
SELECT ENR
FROM PROJECT PT, ASSIGNMENT PM
WHERE DURATION>10 AND BUDGET>1.000.000
AND PT.PNR=PM.PNR
```

Literatur

[1] Erhard Rahm. Mehrrechner-Datenbanksysteme: Grundlagen der verteilten und parallelen Datenbankverarbeitung. Addison-Wesley Bonn, 1994