Task 1 Explain the different phases of query execution.

Task 2 Algebraic Optimization

(a) Which algebraic rules for optimization do you know? Demonstrate their validity with the help of an example respectively.

(b) Given the following tables:

- Persons(PANR, Name, ZIP, Location, Str)
- Phone(PANR, PhoneNr)
- Employees(PANR, PersonnelNr, Department, Salary, Room)
- Students(PANR, MatrNr, FieldOfStudyId)
- Lecture(LID, LName, CH, FieldOfStudyId, Lecturer)
- Registration(LID, MatrNr, Date)

Additionally, the following views exist:

- Phone_directory: SELECT * FROM Persons NATURAL JOIN Phone
- Student_data: SELECT PANR, Name, MatrNr FROM Persons NATURAL JOIN Students
- Exam_list: SELECT * FROM Registration NATURAL JOIN Students NATURAL JOIN Persons

Convert the following queries into relational algebra and apply algebraic optimization to them:

i. SELECT PhoneNr FROM Persons NATURAL JOIN Phone NATURAL JOIN Employees WHERE Name="Paul Dietrich" OR Room="G59-311"

ii. SELECT PhoneNr FROM Student_data NATURAL JOIN Phone_directory WHERE Name like "%Meier" AND FieldOfStudyId=5

iii. SELECT Name FROM Student_data NATURAL JOIN Exam_list NATURAL JOIN Lecture NATURAL JOIN Employees NATURAL JOIN Phone_directory WHERE Name= 'Gunter Saake' AND Lecturer=PersonnelNr
(c) Given the following, additional optimization rule for algebraic optimization:

\[ \pi_x(r_1 - r_2) \leftrightarrow \pi_x(r_1) - \pi_x(r_2) \]

Discuss this rule: Is it correct? For what basic conditions it is applicable?

Task 3 Physical Optimization

Given the following relations:

CUSTOMER(CName, Cadr, IBAN)
ORDER(CName→Customer, Date, Product, Amount)

- Relation Customer: 1000 tuples; one page: 5 tuples
- Relation Order: 100,000 tuples; one page: 10 tuples
- 500 of the orders contain coffee
- Tuples of (CName, IBAN): 50 fit one page
- 3 rows of CUSTOMER × ORDER fit a page
- Buffer: 1 page; no spanned records

Query:

SELECT c.CName, IBAN FROM CUSTOMER c, ORDER o WHERE c.CName = o.CName AND Product = 'coffee'

What is the optimal query plan for these conditions?
How do the needed read-/write accesses differ under the following assumptions?

(a) The relation CUSTOMER is a view. The underlying relation PERSONS has 10000 entries with 2 tuples fitting one page:

CREATE VIEW Customer AS SELECT CName, Cadr, IBAN FROM Persons WHERE purchase > 0;

(b) The table customer is organized as a B+-Tree
(c) There is a hash-index on the product attribute of the order relation
(d) The buffer size is 10 pages
(e) The query is: Product <> 'coffee'

Good Luck!