Exercises
(Course: Database Management Systems)
Chapter 3
Algorithms for Query Processing and Optimization
(Cost-based optimization)

Given a database schema as follows:

```
CREATE TABLE EMPLOYEE
        ( FNAME             VARCHAR(15)         NOT NULL ,
          MINIT             CHAR ,
          LNAME             VARCHAR(15)         NOT NULL ,
          SSN               CHAR(9)             NOT NULL ,
          BDATE             DATE,
          ADDRESS           VARCHAR(30) ,
          SEX               CHAR ,
          SALARY            DECIMAL(10,2) ,
          SUPERSSN          CHAR(9) ,
          DNO               INT                 NOT NULL ,
    PRIMARY KEY (SSN) ,
    FOREIGN KEY (SUPERSSN) REFERENCES EMPLOYEE(SSN) ,
    FOREIGN KEY (DNO) REFERENCES DEPARTMENT(DNUMBER) ) ;

CREATE TABLE DEPARTMENT
        ( DNAME             VARCHAR(15)         NOT NULL ,
          DNUMBER           INT                 NOT NULL ,
          MGRSSN            CHAR(9)             NOT NULL ,
          MGRSTARTDATE      DATE ,
        PRIMARY KEY (DNUMBER) ,
        UNIQUE (DNAME) ,
        FOREIGN KEY (MGRSSN) REFERENCES EMPLOYEE(SSN) ) ;

CREATE TABLE PROJECT
        ( PNAME             VARCHAR(15)         NOT NULL ,
          PNUMBER           INT                 NOT NULL ,
          PLOCATION         VARCHAR(15) ,
          DNUM              INT                 NOT NULL ,
        PRIMARY KEY (PNUMBER) ,
        UNIQUE (PNAME) ,
        FOREIGN KEY (DNUM) REFERENCES DEPARTMENT(DNUMBER) ) ;
```

The **EMPLOYEE** file has $r_E = 10000$ records stored in $b_E = 2000$ disk blocks with the blocking factor $bfr_E = 5$ records/block and the following access paths:
1. A clustering index on SALARY, with levels $x_{SALARY} = 3$ and first-level (base level) index blocks $bl1_{SALARY} = 50$ and average selection cardinality $s_{SALARY} = 20$.
2. A secondary index on the key attribute SSN, with $x_{SSN} = 4$ ($s_{SSN} = 1$) and first-level (base level) index blocks $bl1_{SSN} = 50$.

3. A secondary index on the nonkey attribute DNO, with $x_{DNO} = 2$ and first-level (base level) index blocks $bl1_{DNO} = 4$. There are $d_{DNO} = 125$ distinct values for DNO, so the selection cardinality of DNO is $s_{DNO} = (r_E/d_{DNO}) = 80$.
4. A secondary index on SEX, with $x_{SEX} = 1$. There are $d_{SEX} = 2$ values for the SEX attribute, so the average selection cardinality is $s_{SEX} = r_E*(1/d_{SEX}) = 5000$.

The **DEPARTMENT** file consists of $r_D = 125$ records stored in $b_D = 13$ disk blocks with the blocking factor $bfr_D = 10$ records/block and the following access paths:
1. A primary index on DNUMBER with $x_{DNUMBER} = 1$ level ($s_{DNUMBER} = 1$).
2. A secondary index on MGRSSN with selection cardinality $s_{MGRSSN} = 1$ and levels $x_{MGRSSN} = 2$ and first-level (base level) index blocks $bl1_{MGRSSN} = 2$.

The **PROJECT** file has $r_P = 2000$ records stored in $b_P = 100$ blocks with the blocking factor $bfr_P = 20$ records/block and the following access paths:
1. A secondary index on PLOCATION with $x_{PLOCATION} = 2$ and first-level (base level) index blocks $bl1_{PLOCATION} = 4$. There are $d_{PLOCATION} = 200$ values for the PLOCATION attribute, so the average selection cardinality is $s_{PLOCATION} = r_P*(1/d_{PLOCATION}) = 10$.

Determine the reasonable execution plans for each query below using cost-based optimization:

E1. $\sigma_{SSN = '123456789'}(EMPLOYEE)$

Solution: use secondary index-based search with 5 block accesses.

E2. $\sigma_{DNO=5}(EMPLOYEE)$

Solution: use secondary index-based search with 82 block accesses.

E3. $\sigma_{DNO>5}(EMPLOYEE)$

Solution: use linear search with 2000 block accesses.

E4. $\sigma_{DNUMBER>5}(DEPARTMENT)$

Solution: use primary index-based search for multiple records for the comparison on the key field with 8 block accesses.

E5. $\sigma_{PLOCATION='Stafford'}(PROJECT)$

Solution: use linear search with 100 block accesses.

E6. $\sigma_{DNO=5 \text{ and } SALARY>30000 \text{ and } SEX='F'}(EMPLOYEE)$

Solution: use secondary index-based search on DNO with 82 block accesses and the other conditions are checked later.

E7. $\sigma_{DNO=5 \text{ or } SALARY>30000}$(EMPLOYEE)

Solution: use the union the results from the approach on DNO and from the approach on SALARY with the total block accesses = 1085 block accesses.

E8. EMPLOYEE $\bowtie_{DNO=DNUMBER}$DEPARTMENT

Given: the join selectivity is $js_{E8} = 1/|DEPARTMENT| = 1/r_D = 1/125$ because DNUMBER is a key of DEPARTMENT; the blocking factor for the resulting join file is $bfr_{ED} = 4$ records per block.

Solution: use J2b. Single-loop join with DEPARTMENT for outer loop and a secondary index on the nonkey attribute DNO of EMPLOYEE with the total number of block accesses = 12763 block accesses.

E9. DEPARTMENT $\bowtie_{MGRSSN=SSN}$EMPLOYEE

Given: the join selectivity is $js_{E9} = 1/|EMPLOYEE| = 1/r_E = 1/10000$ because SSN is a key of EMPLOYEE; the blocking factor for the resulting join file is $bfr_{ED} = 4$ records per block.

Solution: use J2b. Single-loop join with DEPARTMENT for outer loop and a secondary index on the key attribute SSN of EMPLOYEE with the total number of block accesses = 670 block accesses.

E10.
```
SELECT *
FROM EMPLOYEE, DEPARTMENT
WHERE DNAME = 'Research' and DNUMBER = DNO;
```

Given: the join selectivity is $js_{E10} = 1/|DEPARTMENT| = 1/r_D = 1/125$ because DNUMBER is a key of DEPARTMENT; the blocking factor for the resulting join file is $bfr_{ED} = 4$ records per block.

Solution: the execution plan for the query is as follows:
- use linear search approach for an equality condition on the unique attribute DNAME to retrieve a single record (in DEPARTMENT*) with the total number of block accesses = 7 block accesses;
- use J2. Single-loop join with DEPARTMENT* for outer loop and a secondary index on the nonkey attribute DNO of EMPLOYEE with the total number of block accesses = 2583 block accesses;
- the total number of block accesses at the end is 7 + 2583 = 2590 block accesses.

E11.
```
SELECT *
FROM EMPLOYEE as E, EMPLOYEE as S
WHERE E.SUPERSSN = S.SSN;
```

Given: the join selectivity is $js_{E11} = 1/|EMPLOYEE| = 1/r_E = 1/10000$ because SSN is a key of EMPLOYEE; the blocking factor for the resulting join file is $bfr_{EE} = 2$ records per block; the available buffer space in main memory is $n_B = 7$ blocks.

Solution: use J3. Sort-merge join with the assumption that the ordering (sorting) of records is obtained via the secondary index on the key attribute SSN of EMPLOYEE as E and external sorting is required for the ordering of records based on values of the nonkey attribute SUPERSSN of EMPLOYEE as S with the total number of block accesses = 39054 block accesses.

E12. SELECT *
FROM PROJECT, DEPARTMENT, EMPLOYEE
WHERE DNUM = DNUMBER and MGRSSN = SSN and LNAME = 'Smith';

Given: the join selectivity with the join condition DNUM = DNUMBER is $js_1 = 1/|DEPARTMENT| = 1/r_D = 1/125$ because DNUMBER is a key of DEPARTMENT; the join selectivity with the join condition MGRSSN = SSN is $js_2 = 1/|EMPLOYEE| = 1/r_E = 1/10000$ because SSN is a key of EMPLOYEE; the blocking factor for the resulting join file as well as for the intermediate join file is $bfr_{join} = 2$ records per block.

Solution: Refer to [1], 15.8.6, pp. 530-532.

E13. SELECT *
FROM PROJECT, DEPARTMENT, EMPLOYEE
WHERE DNUM = DNUMBER and MGRSSN = SSN and PLOCATION = 'Stafford';

Given: the join selectivity with the join condition DNUM = DNUMBER is $js_1 = 1/|DEPARTMENT| = 1/r_D = 1/125$ because DNUMBER is a key of DEPARTMENT; the join selectivity with the join condition MGRSSN = SSN is $js_2 = 1/|EMPLOYEE| = 1/r_E = 1/10000$ because SSN is a key of EMPLOYEE; the blocking factor for the resulting join file as well as for the intermediate join files is $bfr_{join} = 2$ records per block.

Solution: Refer to [1], 15.8.6, pp. 530-532.