Chương 0

Câu 1.Nêu các khái niệm và tính năng về DBMS

Câu 2. Nêu các khả năng của một DBMS

Câu 3. Sơ lược lịch sử phát triển của các DBMS

Câu 4. Nêu các nhiệm vụ và kỹ năng cần thiết của một DBA

Câu 5. Trình bày sơ lược kiến trúc của một DBMS

Câu 6. Giải thích chi tiết các thành phần của một DBMS

Chương 1

Câu 1. Hãy nêu các khác biệt (nếu có) giữa *file organization* và *access method*.

Câu 2. Hãy nêu sự khác nhau giữa *static* và *dynamic files*.

Câu 3. Các tác vụ tập tin nào phụ thuộc vào vị trí bản ghi hiện tại trong tập tin?

Câu 4. Hãy nêu các ưu điểm và nhược điểm khi dùng (a) unordered file, (b) ordered file, và (c) static hash file dùng ith buckets và chaining. Các tác vụ nào thực hiện hiệu quả (không hiệu quả) trên các tổ chức tập tin trên?

Câu 5. Mô tả các kỹ thuật cho phép một tập tin hash tăng và giảm kích thước động. Ưu và nhược điểm của các kỹ thuật này?

Câu 6. Consider a disk with the following characteristics (these are not parameters of any particular disk unit): block size B = 512 bytes; interblock gap size G = 128 bytes; number of blocks per track = 20; number of tracks per surface = 400. A disk pack consists of 15 double-sided disks.

a. What is the total capacity of a track, and what is its useful capacity (excluding interblock gaps)?

b. How many cylinders are there?

c. What are the total capacity and the useful capacity of a cylinder?

d. What are the total capacity and the useful capacity of a disk pack?

e. Suppose that the disk drive rotates the disk pack at a speed of 2400 rpm (revolutions per minute); what are the transfer rate (rr) in bytes/msec and the block transfer time (btt) in msec? What is the average rotational delay (rd) in msec?

What is the bulk transfer rate?

f. Suppose that the average seek time is 30 msec. How much time does it take (on the average) in msec to locate and transfer a single block, given its block address?

g. Calculate the average time it would take to transfer 20 random blocks, and compare this with the time it would take to transfer 20 consecutive blocks using double buffering to save seek time and rotational delay.

Câu 7. A file has r = 20,000 STUDENT records of fixed length. Each record has the following fields: NAME (30 bytes), SSN (9 bytes), ADDRESS (40 bytes), PHDNE (9 bytes), BIRTHDATE (8 bytes), SEX (l byte), MAJORDEPTCODE (4 bytes), MINORDEPTCODE (4 bytes), CLASSCODE (4 bytes, integer), and DEGREEPROGRAM (3 bytes). An additional byte is used as a deletion marker. The file is stored on the disk whose parameters are given in Exercise 6.

a. Calculate the record size R in bytes.

b. Calculate the blocking factor bfr and the number of file blocks b, assuming an unspanned organization.

c. Calculate the average time it takes to find a record by doing a linear search on the file if (i) the file blocks are stored contiguously, and double buffering is used; (ii) the file blocks are not stored contiguously.

d. Assume that the file is ordered by SSN; calculate the time it takes to search for a record given its SSN value, by doing a binary search.

Câu 8. A PARTS file with Part# as hash key includes records with the following Part# values: 2369,3760,4692,4871, 5659, 1821, 1074, 7115, 1620, 2428,3943,4750, 6975, 4981, 9208. The file uses eight buckets, numbered 0 to 7. Each bucket is one disk block and holds two records. Load these records into the file in the given order, using the hash function h(K) = K mod 8. Calculate the average number of block accesses for a random retrieval on Part#.

Câu 9. Load the records of Exercise 8 into expandable hash files based on extendible hashing. Show the structure of the directory at each step, and the global and local depths. Use the hash function h(K) = K mod 128.

Câu 10. Load the records of Exercise 8 into an expandable hash file, using linear hashing. Start with a single disk block, using the hash function $h_0 = K \bmod 2^0$, and show how the file grows and how the hash functions change as the records are inserted. Assume that blocks are split whenever an overflow occurs, and show the value of n at each stage.

Câu 11. Suppose that we have a hash file of fixed-length records, and suppose that overflow is handled by chaining. Outline algorithms for insertion, deletion, and modification of a file record. State any assumptions you make.

Câu 12. Can you think of techniques other than chaining to handle bucket overflow in external hashing?

Câu 13. Write pseudocode for the insertion algorithms for linear hashing and for extendible hashing.

Chương 2

Q1. Question 14.1 in the text book ("Fundamentals of Database Systems- 4th Edition", Elmasri et al., 2004)

Define the following terms: indexing field, primary key field, clustering field, secondary key field, bl.ock anchor, dense index, and nondense (sparse) index.

Q2. Question 14.2 in the text book

What are the differences among primary, secondary, and clustering indexes? How do these differences affect the ways in which these indexes are implemented? Which of the indexes are dense, and which are not?

Q3. Question 14.3 in the text book

Why can we have at most one primary or clustering index on a file, but several secondary indexes?

Q4. Question 14.4 in the text book

How does multilevel indexing improve the efficiency of searching an index file?

Q5. Question 14.5 in the text book

What is the order p of a B-tree? Describe the structure of B-tree nodes.

Q6. Question 14.6 in the text book

What is the order p of a B^+-tree? Describe the structure of both internal and leaf nodes of a B^+-tree.

Q7. Question 14.7 in the text book

How does a B-tree differ from a B^+-tree? Why is a B^+-tree usually preferred as an access structure to a data file?

Exercises

E1. Exercise 14.15in the text book ("Fundamentals of Database Systems- 4th Edition", Elmasri et al., 2004)

A PARTS file with Part# as key field includes records with the following Part# values: 23, 65, 37, 60, 46, 92, 48, 71, 56, 59, 18, 21, 10, 74, 78, 15, 16, 20, 24, 28, 39, 43, 47, 50, 69, 75, 8, 49, 33, 38. Suppose that the search field values are inserted in the given order in a B^+-tree of order p = 4 and p_{leaf} = 3; show how the tree will expand and what the final tree will look like.

E2. Exercise 14.16in the text book

Repeat Exercise 14.15, but use a B-tree of order p = 4 instead of a B^+-tree.

E3. Exercise 14.17in the text book

Suppose that the following search field values are deleted, in the given order, from the B^+-tree of Exercise 14.15; show how the tree will shrink and show the final tree. The deleted values are 65, 75,43, 18,20,92,59,37.

E4. Exercise 14.18in the text book

Repeat Exercise 14.17, but for the B-tree of Exercise 14.16.

Chương 3

Q1. Question 15.1 in the text book("Fundamentals of Database Systems- 4th Edition", Elmasri et al., 2004)

Discuss the reasons for converting SQL queries into relational algebra queriesbefore optimization is done.

Q2. Question 15.3 in the text book

What is a query execution plan?

Q3. Question 15.4 in the text book

What is meant by the term heuristic optimization? Discuss the main heuristics that are applied during query optimization.

Q4. Question 15.5 in the text book

How does a query tree represent a relational algebra expression? What is meant by an execution of a query tree? Discuss the rules for transformation of query trees,and identify when each rule should be applied during optimization.

Q5. Question 15.6 in the text book

How many different join orders are there for a query that joins 10 relations?

Q6. Question 15.7 in the text book

What is meant by cost-based query optimization?

Q7. Question 15.8 in the text book

What is the difference between pipelining and materialization?

Q8. Question 15.9 in the text book

Discuss the cost components for a cost function that is used to estimate query execution cost. Which cost components are used most often as the basis for costfunctions?

Exercises

E1. Exercise 15.13in the text book ("Fundamentals of Database Systems- 4th Edition", Elmasri et al., 2004)

Consider SQL queries Ql, Q8, QIB, Q4, and Q27 from Chapter 8.

```
Q1:  SELECT FNAME,LNAME,ADDRESS
     FROM EMPLOYEE,DEPARTMENT
     WHERE DNAME='Research' AND DNUMBER=DNO;

Q8:  SELECT E.FNAME, E.LNAME, S.FNAME, S.LNAME
     FROM EMPLOYEE AS E, EMPLOYEE AS S
     WHERE E.SUPERSSN=S.SSN;

QIB: SELECT E.FNAME, E.NAME, E.ADDRESS
     FROM EMPLOYEE E, DEPARTMENT D
     WHERE D.NAME='Research' AND D.DNUMBER=E.DNUMBER;

Q4:  (SELECT DISTINCT PNUMBER
     FROM PROJECT, DEPARTMENT, EMPLOYEE
     WHERE DNUM=DNUMBER AND MGRSSN=SSN AND LNAME='Smith')
     UNION
     (SELECT DISTINCT PNUMBER
     FROM PROJECT, WORKS_ON, EMPLOYEE
     WHERE PNUMBER=PNO AND ESSN=SSN AND LNAME='Smith');
```

```
Q27: SELECT PNUMBER, PNAME, COUNT (*)
     FROM PROJECT, WORKS_ON, EMPLOYEE
     WHERE PNUMBER=PNO AND SSN=ESSN AND DNO=5
     GROUP BY PNUMBER, PNAME;
```

a. Draw at least two query trees that can represent each of these queries. Under what circumstances would you use each of your query trees?

b. Draw the initial query tree for each of these queries, then show how the querytree is optimized by the algorithm outlined in Section 15.7.

c. For each query, compare your own query trees of part (a) and the initial andfinal query trees of part (b).

E2. Exercise 15.14in the text book

A file of 4096 blocks is to be sorted with an available buffer space of 64 blocks.How many passes will be needed in the merge phase of the external sort-merge algorithm?

E3.Given the following relations:

EMPLOYEE(ename, ssn, bdate, address, sex, salary, dno)

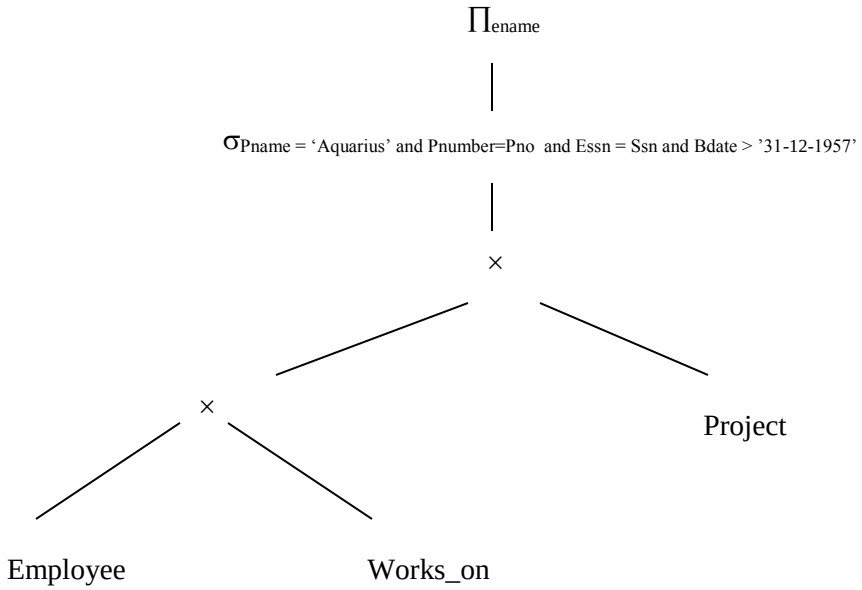
PROJECT(pname, pnumber, plocation)

WORKS_ON(essn, pno, hours)

and the query:

> "Find the names of the employees whose birthdates are after 1957 and currently work for the project *Aquarius*"

Apply the heuristic optimization transformation rules to find an efficient query execution plan for the above query, which is described by the following query tree.



E4

Given the three following relations:

Supplier(Supp#, Name, City, Specialty)

Project(Proj#, Name, City, Budget)

Order(Supp#, Proj#, Part-name, Quantity, Cost)

and the SQL query:

```
SELECT Supplier.Name, Project.Name
FROM Supplier, Order, Project
WHERE Supplier.City = 'New York City' AND Project.Budget > 10000000 AND
    Supplier.Supp# = Order.Supp# AND Order.Proj# = Project.Proj#
```

a. Write the relational algebraic expression that is equivalent to the above query and draw a query tree for the expression.
b. Apply the heuristic optimization transformation rules to find an efficient query execution plan for the above query. Assume that the number of the supliers in New York is larger that the number of the projects with the budgets more than 10000000$.