GATE
2019

DATABASE MANAGEMENT SYSTEM

COMPUTER SCIENCE & INFORMATION TECHNOLOGY

ECG Publications

A Unit of **ENGINEERS CAREER GROUP**

Head O ce: S.C.O-121-122-123, 2nd Floor, Sector-34/A, Chandigarh-160022

Website: www.engineerscareergroup.in **Toll Free:** 1800-270-4242

E-Mail: ecgpublica ons@gmail.com | info@engineerscareergroup.in

GATE-2019: Database Management System | Detailed theory with GATE previous year papers and detailed solu ons.

First Edi on: 2016

Price of Book: *INR 270/-*

CONTENTS

CHAPTER - 1
DATA BASE AND ITS FUNDAMENTALS

1.1 INTRODUCTION

1. Database is a collection of files containing related information.
2. It is managed by popular software systems called database management system (DBMS)

Example: University database contain entities such as students, faculty, courses and classrooms which related to each other by relationship such as student enrollment in courses, faculty, teaching courses.

1.1.1 Features of Database

1. It is logically related
2. It is accessible indifferent orders
3. It is stored only once

1.2 TERMINOLOGY OF DATABASE SYSTEM

1. Data

It can be defined as representation of facts, concepts or instruction in a formalized manner suitable for interpretation or processing by human or electronic machine. It can be represented with the help of characters like alphabet (A – 2, a – 3) digits (0 – 9) or special characters (+, –, %, <, > etc)

(i) Data Item (File)

It is a set of character which is used together to represent specific data element e.g. name of student can be represented by NAME.

(ii) Record

A record is a collection of related data items e.g. a payroll record of employee, profile of a student in college.

(iii) File (Data File)

File is a collection of related record e.g. a payroll file might consist of the employees pay records for a company.

For example:

Roll No.	Name	Marks
101	Rachit	85
105	Rohit	75

Here, given table is a file and Roll No., Name, Marks are fields.

(iv) Information

Information is organized or classified data and it has some meaningful value.

Information has following characteristics:

(a) Timely
(b) Accurate
(c) Complete
(d) Given to right person

ASSIGNMENT - I

1. Find the number of candidate keys of relation R (A, B, C, D, E, F, G, H, I, J) with F.D set ABD→E, C→J, AB→G, CI→J, B→F, G→H.

2. How many super keys are possible with candidate by having m attributes if original relation has T member of attributes?

3. Find the Prime attributes in relation R(A, B, C, D, E) with F.D set = { AB→C, C→D, D→E, C→A}.

4. How many total candidate keys are possible for relation R(A, B, C, D) with F.D set = {A→B, B→ C, C →D, D→A}?

5. Find the candidate keys of relation R(A, B, C, G, H, I} with F.D set {A→B, A →C, CG→A, CG→I, B→H}.

6. Find number of candidate keys of relation R(A, B, C, D, E, F) with F.D set = {A→C, B→D, C →E, D→E, E→A, F→B}.

7. Find the set of decomposed relation R(A, C, D) using F.D set = { AB→C, C→A, C→D,} of relation R(A, B, C, D).

8. Find the candidate keys of relation R(A, B, C, D, E, F) having F.D set {AB→CD, C→D, D→E, E→F}.

9. Check whether decomposition D = {{A, B, C}, {A, C, D, E}} of relation R (ABCDE) is Lossless or Lossy if F.D set of R is {A → BC, CD → E, B →D, E → A}

10. Check whether decomposition set D = {ABC, CD, DE} is dependency preserving or not for original relation R(ABCDE) with F.D set = {AB → CD, C → D, D → E}

11. Consider a relation R(A, B, C, D, E, F) with F.D set F = {AB → CD, C → D, D → E, E → F} and decomposition set D = {AB, CDE, EF}. Check whether it is dependency preserving or not.

12. What is the Highest Normal Form of following relations

(a) R (ABCD) with F.D set = {AB → C, BC → D}

(b) R(A, B, C, D, F, E) with F.D set {AB → C, C → DE, E → F, D → A}

(c) R(ABCDEFGH) with F.D set {AB → CD, D → EG, F → H, C → EF, H → A, G → B, A → B}

13. Check the Highest normal form of relation R(A, B, C, D) with F.D set {A → B, B → C, C → D, D → A} and decompose it, if it is not in BCNF.

14. What is the HNF of the Relation R (ABCDEFGHI) with F.D set = {AB → CD, C → DF, E → GH, C → E, H → I, F → B}.

15. Consider the relation schema EMP_DEPT with following set G of functional dependencies G = {ENO → {ENAME, DOB, ADDRESS, DNUMBER}, DNUMBER → {DNAME, DMGRENO}}. Calculate the closures of $\{ENO\}^+$ and $\{DNUMBER\}^+$ with respect to G.

16. Consider the relation R, which has attributes that hold schedules of courses and sections t an university; R = {CourseNo, SecNo, OfferingDept, Credit-Hours, CourseLevel, InstructorENO, Semester, Year, Days_Hours, RoomNo, NoOfStudents}. Suppose that the following functional dependencies hold on R:

{CourseNo} → {OfferingDept, CreditHours, CourseLevel}

{CourseNo, SecNo, Semester, Year} → {InstructroENO, CourseNo, SecNo}

Try to determine which sets of attributes form keys of R. How would you normalize this relation?

SOLUTIONS

Sol. 1.
R(A, B, C, D, E, F, G, H, I, J)
ABD → E, C → J, AB → G, CI → J, B → F, G → H.
ABD^+ = ABDEGFH
C^+ = CJ
AB^+ = ABGH
CI^+ = CIJ
B^+ = BF
G^+ = GH
Actually there is no candidate key ,But we have to make any of the above closure as candidate key in which we have to do less inclusion of attribute. So, $ABCDI^+$ will be candidate key.

Sol. 2.
2^{T-m}

Sol. 3.
R(ABCDE)
FD = { AB → C, C → D, D → E, C → A}
AB^+ = ABCDE CB^+ = ABCDE
C^+ = CDEA
AB and CB are candidate keys. So, A, B and C are prime attributes

Sol. 4.
R (A, B, C, D)
A^+ = ABCD
C^+ = ABCD
C^+ = ABCD
D^+ = ABCD
So total candidate keys = 4

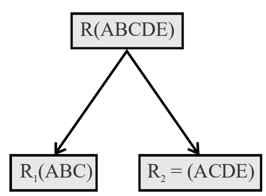
Sol. 5.
Its candidate key is CG only as its closure
$(CG)^+$ = {CGAIBH}

Sol. 6.
R (A, B, C, D, E, F)
F.D = { A → C, B → D, C → E, D → E, E → A, F →B}
A^+ = ACE
B^+ = BDEAC
C^+ = ACE
D^+ = DEAC
E^+ = EAC
C^+ = FBDEAC = ABCDEF
Number of Candidate keys = 1.

Sol. 7.
R(A, C, D) is decomposed relation of R(A, B, C, D)
So Functional dependency will become as
F = { C→ A, C → D}

Sol. 8.
R = (A, B, C, D, E, F)
{AB → CD, C → D, D → E, E → F}
AB^+ = ABCDEF
C^+ = CDEF
D^+ = DEF
E^+ = EF
∴Candidate key is AB.

Sol. 9.



1. $R_1 \cup R_2$= R
2. $R_1 \cap R_2$= AC
3. $\langle AC \rangle^+$ = ABCDE

This will act key for both relations. So, It is a lossless decomposition.

GATE QUESTIONS

1. Consider the following tables T_1 and T_2.

T_1	
P	**Q**
2	2
3	8
7	3
5	8
6	9
8	5
9	8

T_2	
R	**S**
2	2
8	3
3	2
9	7
5	7
7	2

In table T_1, P is the primary key and Q is the foreign key reference R in table T_2 with on-delete cascade and on-update cascade. In table T_2, R is the primary key and S is the foreign key referencing P in table T_1 with on-delete set NULL and on-update cascade. In order to delete record (3, 8) from table T_1, the number of additional records that need to be deleted from table T_1 is ____________.

[GATE - 2017]

2. An ER model of a database consists of entity types A and B. These are connected by a relationship R which does not have its own attribute. Under which of the following conditions, can the relational table for R be merged with that of A?

[GATE - 2017]

(a) Relationship R is one-to-many and the participation of A in R is total
(b) Relationship R is one-to-many and the participation of A in R is partial
(c) Relationship R is many-to-one and the participation of A in R is total
(d) Relationship R is many-to-one and the participation of A in R is partial

3. Which of the following is NOT a superkey in a relational schema with attributes V, W, X, Y, Z and primary key V Y?

[GATE - 2016]

(a) VXYZ (b) VWXZ
(c) VWXY (d) VWXYZ

4. A database of research articles in a journal uses the following schema.
(VOLUME, NUMBER, STARTPAGE, ENDPAGE, TITLE, YEAR, PRICE)
The primary key is (VOLUME, NUMBER, STARTPAGE, ENDPAGE) and the following functional dependencies exist in the schema.
(VOLUME, NUMBER, STARTPAGE, ENDPAGE)→ TITLE
(VOLUME, NUMBER) → YEAR
(VOLUME, NUMBER, STARTPAGE, ENDPAGE)→ PRICE
The database is redesigned to use the following schemas.
(VOLUME, NUMBER, STARTPAGE, ENDPAGE, TITLE, PRICE)
(VOLUME, NUMBER, YEAR)
Which is the weakest normal form that the new database satisfies, but the old one does not?

[GATE - 2016]

(a) 1NF (b) 2NF
(c) 3NF (d) BCNF

5. Consider an Entity-Relationship (ER) model in which entity sets E1 and E2 are connected by an m : n relationship R12, E1 and E3 are connected by a 1 : n (1 on the side of E1 and n on the side of E3) relationship R13. E1 has two single-valued attributes a11 and a12 of which a11 is the key attribute. E2 has two single-valued attributes a21 and a22 of which a21 is the key attribute. E3 has two single-valued attributes a31 and a32 of which a31 is the key attribute. The relationships do not have Key attributes. The relationship do not have any attribute. If a relational model is derived from the above ER model, then the minimum number of relations that would be generated if all the relations are in 3NF is ________.

[GATE - 2015]

CHAPTER - 2
TRANSACTION PROCESSING CONCEPTS

2.1 INTRODUCTION
A database system is classified on the basis of number of users who can use the system at the same time.
DBMS is classified into
1. Single user DBMS
2. Multiple user DMBS

2.1.1 Single user DBMS
A DBMS is single-user if at most one user at a time can use the system. Single user DBMS are mostly restricted to personal computer system.

2.1.2 Multi user DBMS
1. A DBMS is Multi-user if at most one user at a time(concurrency) can use the system.
2. Concurrency can be achieved either using single processor or multiprocessors.
3. In Single processor, concurrency is achieved using interleaving technique while multiprocessors use parallel processing technique for concurrency.

Example.
(i) An Airline reservation system that is used by number of travel agents and reservation clerks concurrently.
(ii) Online Banking System where number of transactions are processed of customers concurrently.

2.2 TRANSACTION
2.2.1 Two important assumptions about transaction:
1. Transaction interacts with each other only via database read and write operations.
2. A database is fixed collection of independent objects. When objects are added to or deleted from a database or there are relationships between database objects that we want to exploit for performance, some additional issue arise.
3. A transaction is an executing program that forms a logical unit of database processing. It is an atomic unit of work that is either completed entirely or not done at all.
4. A transaction includes one or more database access operations (insertion, deletion, modification, or retrieval).
5. The transactions boundaries are specified by explicit begin and end statements.
6. If the database operations in a transaction do not update the database but only retrieve data, the transaction is called a read-only transaction.
7. Basic database access operations of transactions are
(i) Read_ item (x): It reads a database item named X. It can be written as r(x).
(ii) Write_item(x): It writes a database item named X. It can be written as w(x).
8. A transaction includes read and write operations to access and update the database.
9. Execution of read operation command
10. Find the address of the disk block that contains items X.
11. Copy that disk block into a buffer in main memory (if that disk block is not already in some main memory buffer).
12. Copy item X from the buffer to the program variable named X.

ASSIGNMENT

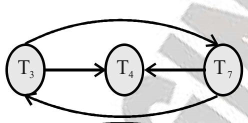
1. The concept of locking can be used to solve the problem of
(a) Lost update
(b) Uncommitted dependency
(c) Inconsistent data
(d) Deadlock

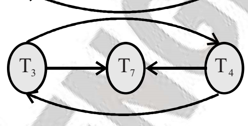
2. What are the potential problems when a DBMS executes multiple transitions concurrently?
(i) The lost update problem
(ii) The dirty read problem
(iii) The unrepeatable problem
(iv) The phantom problem
(a) 3 and 4 only
(b) 1, 2 and 4 only
(c) 2 and 3 only
(d) All of 1, 2, 3 and 4

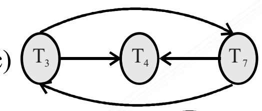
3. Assume transaction A holds a shared lock R. If transaction B also request for a shared lock on R, it will
(a) Result in a deadlock situation
(b) Immediately be granted
(c) Immediately be rejected
(d) Be granted as soon as it is released by A

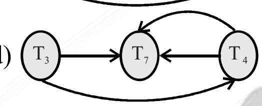
4. Consider the following transaction schedule

T_3	T_4	T_7
R(Q)		
	W(Q)	
W(Q)		R(Q)
		W(Q)

(a) 

(b) 

(c) 

(d) 

5. Consider the following three schedules of transactions T1, T2 and T3. [Notation: In the following NYO represent the action Y (Y or read, write) performed by transaction N on object O.]

(S1)	(S2)	(S3)
2RA	3RC	2RA
2WA	2RA	3RC
3RC	2WA	3WA
2WB	2WB	2WA
3WA	3WA	2WB
3WC	1RB	3WC
1RA	1RB	1RA
1RB	1WA	1RB
1WA	1WB	1WA
1WB	3WC	1WB

Which of the following statements is TRUE?
(a) S1, S2 and S3 are all conflict equivalent to each other
(b) No two of S1, S2 and S3 are conflict equivalent to each other
(c) S2 is conflict equivalent to S3, but not to S1
(d) S1 is conflict equivalent to S2 but not to S3

6. In case of timestamp ordering R-timestamp (Q) denotes
(a) The largest timestamp of any transaction that execute read (Q) successfully.
(b) The average timestamp of any transaction that execute read (Q) successfully
(c) The average timestamp of any transaction that execute read (Q) unsuccessfully

(d) The smallest timestamp of any transaction that execute read (Q) successfully

7. Locking was introduced into databases so that
(a) Keys can be provided to maintain security
(b) All simultaneous transactions are prevented
(c) Passwords can be provided to maintain security
(d) Consistency can be enforced.

8. Which level of locking provides the highest degree of concurrency flight in a relational database?
(a) Page
(b) Table
(c) Row
(d) Page, table and row level locking allow the same degree of concurrency

9. Which of the following is true for two-phase locking?
(a) Lock acquisition is the second phase
(b) Locks can be acquired at any time
(c) Locks are acquired in the first phase
(d) None of the above

10. Consider the following statements
(a) S1 : Entire database cannot be locked
(b) S2 : Entire relation can be locked
Which of the above statements is/are true?
(a) S1 only
(b) S2 only
(c) both S1 and S2 are true
(d) both S1 and S2 are false

11. Choose the false statement.
(a) Timestamp protocol is deadlock free
(b) Two phase locking guarantees serializability
(c) Strict two phase locking is deadlock free
(d) Timestamp protocol may not result recoverable schedule

12. In DBMS without concurrency control, what consistency problem does the following transaction schedule depict?

Time	Transaction A	Transaction B
T1	read R	
T2		
T3		Read R
T4		write R

(a) Dirty Read
(b) Uncommitted Dependency
(c) Inconsistent Analysis
(d) Lost Update

13. For the schedule given below, which of the following is correct?
(i) Read A
(ii) Read B
(iii) Write A
(iv) Read A
(v) Write A
(vi) Write B
(vii) Read B
(viii) Write B

(a) This schedule is serializable and can occur in a scheme using 2PL protocol
(b) This schedule is serializable but cannot occur in a scheme using 2PL protocol
(c) This schedule is not serializable but can occur in a scheme using 2PL protocol
(d) This schedule is not serialzable and cannot occur in a scheme using 2PL protocol

14. When n transactions are run concurrently and in an interleaved manner, the number of possible schedule are________.
(a) Much larger than n!
(b) Much lower than n!
(c) Much larger than (n –1)!
(d) Much lower than (n – 1)!

15. Consider the following schedules involving two transactions.
Which one of the following statements is TRUE?
$S_1 : r_1 (X); r_1 (Y); r_2 (X); r_2 (Y); w_2 (Y); w_1 (X)$
$S_2 : r_1 (X); r_2 (X); w_2 (Y); w_2 (Y); r_1 (Y); w_1 (X)$
(a) Both S1 and S2 are conflict serializable
(b) S1 is conflict serializable and S2 is not conflict serializable
(c) S1 is not conflict serialization and S2 is conflict serializable

GATE QUESTIONS

1. Two transactions T_1 and T_2 are given as
$T_1 : r_1(X)w_1(X)r_1(Y)w_1(Y)$
$T_2 : r_2(Y)w_2(Y)r_2(Z)w_2(Z)$
where $r_1(V)$ denotes a read operation by transaction T_1 on a variable V and $w_1(V)$ denotes a write operation by transaction T_1 on a variable V. The total number of conflict serializable schedules that can be formed by T_1 and T_2 is _____________.

[GATE - 2017]

2. Consider the following database schedule with two transactions, T1 and T2.
$S = r_2(X); r_1(X); r_2(Y); w_1(X); r_1(Y); w_2(X); a_1; a_2$
where $r_i(Z)$ denotes a read operation by transaction T_i on a variable Z, $w_i(Z)$ denotes a write operation by T_i on a variable Z and a_i denotes an abort by transaction T_i.
Which one of the following statements about the above schedule is TRUE?

[GATE - 2016]

(a) S is non-recoverable
(b) S is recoverable, but has a cascading abort
(c) S does not have a cascading abort
(d) S is strict

3. Suppose a database schedule S involves transactions $T_1, ..., T_n$. Construct the precedence graph of S with vertices representing the transactions and edges representing the conflicts. If S is serializable, which one of the following orderings of the vertices of the precedence graph is guaranteed to yield a serial schedule?

[GATE - 2016]

(a) Topological order
(b) Depth-first order
(c) Breadth-first order
(d) Ascending order of transaction indices

4. Consider the following two phase locking protocol. Suppose a transaction T accesses (for read or write operations), a certain set of objects $\{O_1,…., O_k\}$. This is done in the following manner.
Step 1: T acquires exclusive locks to $O_1,….., O_k$ in increasing order of their addresses.
Step 2: The required operations are performed
Step 3: All locks are released.
This protocol will

[GATE - 2016]

(a) Guarantee serializability and deadlock-freedom
(b) Guarantee neither serializability nor deadlock-freedom
(c) Guarantee serializability but not dadlock-freedom
(d) Guarantee deadlock-freedom but not serializability

5. Which one of the following is not a part of the ACID properties of database transactions?

[GATE - 2016]

(a) Atomicity
(b) Consistency
(c) Isolation
(d) Deadlock-freedom

6. Consider the following transaction involving two bank accounts x and y.
read (x) ; x : = x – 50; write (x) ; read (y) ; y : = y + 50; write (y)
The constraint that the sum of the accounts x and y should remain constant is that of

[GATE - 2015]

(a) Atomicity (b) Consistency
(c) Isolation (d) Durability

7. Consider a simple check pointing protocol and the following set of operations in the log.
(start, T4); (write, T4, y, 2, 3); (start, T1); (commit, T4); (write, T1, z, 5, 7);
(checkpoint);
(start, T2); (write, T2, x, 1, 9); (commit, T2); (start, T3); (write, T3, z, 7, 2);

SOLUTIONS

Sol 1. (54)
There is only one conflict serializable schedule as
$T_1 \rightarrow T_2$, because last operation of T_1 and first operation of T_2 conflicts each other.
Number of schedules that are conflict serializable to $T_2 \rightarrow T_1$ is 53.
Proof: The operations of T_1 is
_ $R_1(x)$_$W_1(x)$_$R_1(y)$_$W_1(y)$
The first operation of T_2 that conflicts with operation of T_1 is W_2 (y) but not R_2 (z), W_2(z).
The number of places where W_2(y) can appear is
Case-1. $\underline{W_2(y)}$ $R_1(x)$ W_1 (x) $R_1(y)$ W_1 (y)
Case-2. $R_1(x)$ $\underline{W_2(y)}$ W_1 (x) $R_1(y)$ W_1 (y)
Case-3. $R_1(x)$ W_1 (x) $\underline{W_2(y)}$ $R_1(y)$ W_1 (y)

Case-1. The number of positions that $R_2(z)$ $W_2(z)$ can come before $W_2(y)$ is $^5C_1 + {^5C_2} = 15$ (either both can take same space or two different spaces).
Case-2. The number of positions that $R_2(z)$ $W_2(z)$ can come before $W_2(y)$ is $^4C_1 + {^4C_2} = 10$
For each of these 10 positions $R_2(y)$ can take 2 positions before $W_2(y)$ therefore total possible schedules are $10 \times 2 = 20$
Case-3. The number of positions that $R_2(z)$ $W_2(z)$ can come before $W_2(y)$ is $^3C_1 + {^3C_2} = 6$
For each of these 6 positions $R_2(y)$ can take 3 positions before $W_2(y)$ therefore total possible schedules are $6 \times 3 = 18$.
The total conflict serializable schedules as $T_2 \rightarrow T_1 = 15 + 20 + 18 = 53$
$\therefore$ Total conflict serializable schedules
= 1 + 53 = 54

Sol 2. (c)
As there is no dirty-read in the given schedule, the schedule is both recoverable and cascadeless.

Sol 3. (a)
If a schedule is serializable, the topological order of a graph (precedence graph) yields a serial schedule.

Sol 4. (a)
Two phase locking protocol ensures serializability, but does not ensures freedom from deadlock.

Sol 5. (d)
ACID properties of database transaction is defined as:
A: Atomicity
C: Consistency
I: Isolation
D: Durability

Sol 6. (b)
x and y are bank accounts
Read(x) ; x = x -50
Write(x) ; read(y) ; y = y + 50 ; write(y)
It is the property of consistency directly as it says that sum should remain constant before and after the transaction.
For example, A = 1000 B = 2000
and we want to transfer 500 from A to B.
then A = 500, B = 2500
But sum before and after will remain same.

Sol 7. (a)
Need to undo T_3 and T_1 as they are not commited and redo only T_2

Sol 8. (b)
Schedule is non-recoverable because transaction T_2 commits before T_1 gets failed. So it is non-recoverable even if T_1 go to initial state T_2 can not go because it has committed and committed transaction cannot go back to original position.

Sol 9. (d)

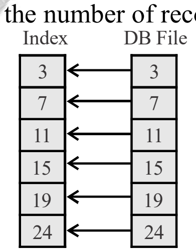
CHAPTER - 3
INDEXING

3.1 INTRODUCTION
1. It is a map to locate records of the database file on disk storage space.
2. It is used to speed up the retrieval of records in response to certain search conditions.
3. It provides secondary access paths, which provide alternative ways of accessing to records without affecting the physical placement of records on disk.
4. To find a record or records in database file, initially index or index file is accessed to get the block address of the search records.
5. Any field of the database file stored on disk can be used to create an index. This field is called indexing field.
6. Any file can have multiple indexes on its different fields.
7. Every index file contains two fields (Searching value, Block pointer) in each entry.

3.2 REQUIREMENT OF INDEXING
The data is shifted from the secondary memory (disk) to the main memory block by block. In the worst case, whole database file will have to be transferred to the main memory. And if binary search is applied for searching any record, $\log_2 N$ blocks will need to be transferred, if file is in N-blocks. So, to minimize the number of blocks to transfer from secondary memory to main memory, indexing is used because index file has very small size.

3.2.1 Important points of Indexing
1. It reduces I/O cost
2. It provides alternative path to access records without affecting the physical records on the disk
3. It is used to step access the desired data.
4. Any field can used create index.
5. Multiple index can exist for same file.
6. Self of attributes which is used to look up records in a file.
7. Index is classified into two categories
(i) Dense Index
(ii) Sparse Index

3.2.2 Dense Index
1. It contains block address of each record of database file.
2. It contains number of entries equal to the number of records in database file.



Example.
If database file contains records of students according to their names in alphabetical order
But its Dense index file contains field student_id to locate records as follows

ASSIGNMENT

1. Suppose that the search field is a non-ordering key field, and we construct a B-tree on this field. Where search field is of 9 bytes, the disk block size is 512 bytes, a record pointer is of 7 bytes and a block pointer is of 6 bytes. Assume each node of the B-tree is 69 percent full. Then calculate the numbers of nodes in this B-tree of level 4.

2. Consider a disk with block size B = 512 bytes. A block pointer is P =6 bytes long, and a record pointer is P_R = 7 bytes long. A file has r = 30,000 EMPLOYEE records of fixed length. Each record has PHONE (9 bytes), DOB (8 bytes), ENO (9 bytes), DEPARTMENTCODE (9 bytes), ADDRESS (40 bytes), PHONE (9 bytes), DOB (8 bytes), SEX (1 bytes), JOBCODE (4 bytes), SALARY (4 bytes, real number). An additional byte is used as a deletion marker:
(i) Calculate the record size R in bytes.
(ii) Calculate the block factor and the number of file blocks b, assuming an un-spanned organization.
(iii) Suppose that the file is ordered by the key field ENO and we want to construct a primary index on ENO. Calculate:
(a) The index blocking factor bfr_i (which is also the index fan-out fo);
(b) The number of first-level index entries and the number of first-level index blocks;
(c) The number of levels needed if we make it into a multilevel index;
(d) The total number of blocks required by the multilevel index; and
(e) The number of block accesses needed to search for and retrieve a record from the file-given its ENO value-using the primary index.

3. Consider a B^+ tree with fan out (the number of block pointers per node) equal to 3 for the following set of key values 80, 50, 10, 70, 30, 100, 90. Assume that the tree is initially empty and the values are added in the order given.
(i) Show the tree after insertion of 10, 30, and 90.
(ii) The key values 30 and 10 are now deleted from the tree in the order. Show the tree after each deletion.

4. A parts file with Part# as key field includes records with the following Part# values: 8, 5, 1, 7, 3, 12, 9, 6. Suppose that the search field values are inserted in the given order in a B^+ tree with leaf order p_{leaf} = 3; show the final tree will-look-like.

SOLUTIONS

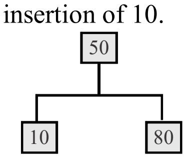
Sol 1.
Apply B-tree structure formula
P(B.P) + (P – l) key + (P – 1) R.P ≤ 512
Let order of B-Tree is P
Given, B.P= 6 bytes, Key = 9 bytes and R.P=7 bytes.
6P + (P-1)(9+7) ≤512
⇒ P = 24
As B-Tree node is 69 % full ∴ P= 0.69*24=17
So, each B-Tree node has 17 pointers and 16 keys

Level	Nodes
1	1
2	17
3	17*17
4	17*17*17

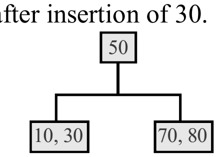
At level 4 , there are 4913 nodes

Sol 2.
(i) Record size R
= 9+8+9+9+40+9+8+1+4+4+1
= 102 Byte
(ii) Block factor = Number of records in one block.
Given,
Block size = 512 bytes
Each record is of size = 102 Bytes
∴ Number of records in a Block = $\frac{512}{102} = 5$ records.
⇒ Block factor = 5
No. of records in file = 30,000
∴ Number of blocks for a file = $\frac{30000}{5} = 6000$ blocks
(a) Entry of primary index file contains key and Block pointer only
∴ Record size of Primary index file = 9 + 6 = 15 bytes
∴ Index blocking factor bfr_i
$= \frac{\text{Block size}}{\text{Size of index file's Record}} = \frac{512}{15} = 34$
(b) No. of entries in Primary index = 30,000
Each block can store 34 entries of primary index.
∴ Number of blocks at the first level of primary index = $\frac{30,000}{34} = 883s$
(c) Number of blocks at Ist level = $\frac{30000}{34} = 883$
Number of blocks at 2[nd] level = $\frac{883}{34} = 26$
Number of blocks at 3[rd] level = $\frac{26}{34} = 1$ block.
∴ There are 3 levels of multilevel Primary index.
(d) Total number of blocks required for primary index = 883 + 26 + 1 = 910
(e) Number of blocks needed to search or retrieve a record from file includes 3 index blocks and 1 data block
∴ Number of blocks required to access a record = 4 blocks

Sol 3.
(i) B^+ tree after insertion of 10.



(ii) B^+ tree after insertion of 30.



(iii) B^+ tree after insertion of 90.

GATE QUESTIONS

1. In a B^+ tree, if the search-key value is 8 bytes long, the block size is 512 bytes and the block pointer size is 2 bytes, then the maximum order of the B^+ tree is __________.

[GATE - 2017]

2. B^+ Trees are considered BALANCED because

[GATE - 2016]

(a) The lengths of the paths from the root to all leaf nodes are all equal.
(b) The lengths of the paths from the root to all leaf nodes differ from each other by at most 1.
(c) The number of children of any two non-leaf sibling nodes differ by at most 1.
(d) The number of records in any two leaf nodes differ by at most 1.

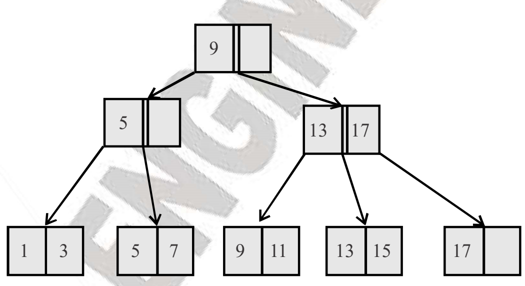
3. A file is organized so that the ordering of data records is the same or close to the ordering of data entries in some index. Then that index is called

[GATE - 2015]

(a) Dense (b) Sparse
(c) Clustered (d) Unclustered

4. With reference to the B+ tree index of order 1 shown below, the minimum number of nodes (including the Root node) that must be fetched in order to satisfy the following query "Get all records with a search key greater than or equal to 7 and less than 15" is ______.

[GATE - 2015]



5. In a B+ tree in which the search key is 12B long, block size is 1024B, record pointer is 10B long and block pointer is 8B long. The maximum number of keys that can be accommodate in each non-leaf node of the tree is _______.

[GATE - 2015]

6. An index is clustered, if

[GATE - 2013]

(a) It is on a set of fields that form a candidate key.
(b) It is on a set of fields that include the primary key.
(c) The data records of the file are organized in the same order as the data entries of the index.
(d) The data records of the file are organized not in the same order as the data entries of the index.

7. Consider a B^+ - tree in which the maximum number of keys in a node is 5. What is the minimum number of keys in any non-root node?

[GATE - 2010]

(a) 1 (b) 2
(c) 3 (d) 4

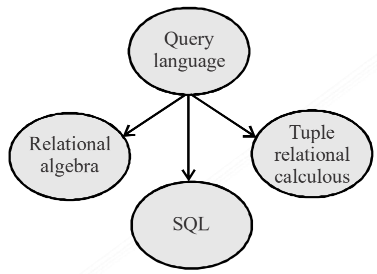
8. The following key values are inserted into a B^+ - tree in which order of the internal nodes is 3, and that of the leaf nodes is 2, in the sequence given below. The order of internal nodes is the maximum number of tree pointers in each node, and the order of leaf nodes is the maximum number of data items that can be stored in it. The B^+- tree is initially empty.
10, 3, 6, 8 , 4, 2 1
The maximum number of times leaf nodes would get split up as a result of these insertions is

[GATE - 2009]

(a) 2 (b) 3
(c) 4 (d) 5

CHAPTER - 4
QUERY LANGUAGE

4.1 INTRODUCTION

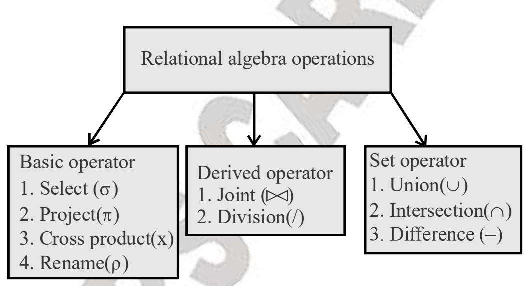


Query is always executed tuple by tuple and one tuple at a time.

4.2 RELATIONAL ALGEBRA (RA)

The relational algebra is a procedural query language. It consists of set of operators that take one or two relations as input and product a new relation as their result. RA forms the core component of a relational query engine. RA provides a framework for query optimization. SQL queries are internally translated into RA expressions.
Relation algebra by default eliminates the duplicate tuples from the resuit.

4.2.1 RA Operations



4.2.1.1 Basic Operator

(i) The Select Operation (σ)

It is a unary operator and is used to select those tuples of a relation that satisfy a given condition.

(a) Notation

σ_θ (r)

Where σ is select operator (sigma)

θ is selection condition, r is relation

(b) Result

A relation with the same schema as r consisting of the tuples in r that satisfy condition θ.

(c) Properties

It is commutative as $\sigma_{C1} (\sigma_{C2} (r)) = \sigma_{C2} (\sigma_{C1} (r))$.

(d) Select Condition

Atomic or composite condition. Composite condition is atomic conditions combined with logical operators AND, OR and NOT.

ASSIGNMENT

1. Which of the following operations is not part of the five basic set operations in relational algebra?
(a) Union (b) Division
(c) Cartesian Product (d) Set Difference

2. Which of the following relational algebraic operation is not a commutative operation?
(a) Union (b) Intersection
(c) Selection (d) Projection

3. Which of the following is wrong?
(a) $\pi L1 \cup L2$ (E1 $\bowtie \theta$ E2) = $(\pi_{L_1}(E_1)) \bowtie \theta$ $(\pi_{L_1}(E_2))$
(b) σP (E1 – E2) = σP (E1) σP (E2)
(c) $\sigma_{\theta_1 \wedge \theta_2}$ (E) = $\sigma_{\theta_1}(\sigma_{\theta_2}(E))$
(d) E1 $\bowtie \theta$ E2 = E2 $\bowtie \theta_1$ E1

4. Which of the following is correct?
(a) An SQL query automatically eliminates duplicates
(b) An SQL query will not work if there are no indexes on the relations
(c) SQL permits attribute names to be repeated in the same relation
(d) None of the above

5. In SQL, relations can contain null values, and comparisons with null values are treated as unknown. Suppose all value are treated as false. Which of the following pairs is not equivalent?
(a) x = 5 not(not(x = 5)
(b) x = 5 x > 4 and x < 6, where x is an integer
(c) x ≠ 5 not (x = 5)
(d) None of the above

6. Consider the following statements:
S1: we can use IN in place of = ANY
S2: we can use IN in place of = ALL
Which one of the following is true?
(a) S1 is true
(b) S2 is true
(c) Both S1 and S2 are true
(d) None of the above

7. Constraints are specified as a part of
(a) Data definition
(b) Data manipulation
(c) Data control
(d) None of the above

8. Which of the following tuple relational calculus finds all customers who have a loan amount of more than 1200?
(a) (t(Customer-name)|t ε borrow Λ t[amount] > 1200}
(b) {t|t (Customer-name) ε borrow Λ t[amount] > 1200}
(c) {t | ∃ s ε borrow (t [Customer_name)| = s [Customer_name] Λ s [amount] > 1200}
(d) None of the above

9. The following tables gives details of employees in a company department

Emp_ID	Job	Salary	Des-ID
110	Designing	25000	SW
115	Calibrating	19000	QA
120	Programming	26000	SW
135	Quality Assuring	18000	QA
150	Consulting	45000	CO
168	Consulting	35000	CO
188	Analysis and Design	22000	SW

Which of the following SQL statement gives the average Salary for each designation ID (Des_ID)?
(a) SELECT AVG (Salary) FROM Employee
(b) SELECT Des-ID, AVG (Salary) FROM Employee ORDER BY Des_ID
(b) SELECT Des_ID, Salary FROM Employee GROUP By Des-ID
(c) SELECT Des_ID, AVG (Salary FROM Employee GROUP BY Des_ID)

GATE QUESTIONS

1. Consider a database that has the relation schema EMP (EmpID, EmpName, and DeptName). An instance of the schema EMP and a SQL query on it are given below.

EMP		
EmpID	**EmpName**	**DeptName**
1	XYA	AA
2	XYB	AA
3	XYC	AA
4	XYD	AA
5	XYE	AB
6	XYF	AB
7	XYG	AB
8	XYH	AC
9	XYI	AC
10	XYJ	AC
11	XYK	AD
12	XYL	AD
13	XYM	AE

```
SELECT AVG(EC.Num)
FROM EC
WHERE (DeptName, Num) IN
    (SELECT DeptName, COUNT (EmpId) AS
                                EC(DeptName,
Num)
    FROM EMP
    GROUP BY DeptName
```

The output of executing the SQL query is ___________.

[GATE - 2017]

2. Consider the following database table named top-scorer.

top_scorer		
player	**country**	**goals**
Klose	Germany	16
Ronaldo	Brazil	15
G Muller	Germany	14
Fontaine	France	13
Pele	Brazil	12
Klinsmann	Germany	11
Kocsis	Hungary	11
Batistuta	Argentina	10
Cubillas	Peru	10
Lato	Poland	10
Lineker	England	10
T Muller	Germany	10
Rahn	Germany	10

Consider the following SQL query:

```
SELECT ta.player FROM top_scorer AS ta
WHERE ta.goals > ALL (SELECT tb.goals)
            FROM top-scorer AS tb
            WHERE tb.country = 'Spain')
AND ta.goals > ANY (SELECT tc.goals FROM
top-scorer AS tc
            WHERE tc.country = 'Germany')
```

The number of tuples returned by the above SQL query is _________.

[GATE - 2017]

3. Consider the following database table named water_schemes:

Water_schemes		
Scheme_no	**District_name**	**Capacity**
1	Ajmer	20
1	Bikaner	10
2	Bikaner	10
3	Bikaner	20
1	Churu	10
2	Churu	20
1	Dungargarg	10

The number of tuples returned by the following SQL query is________.

```
with total(name, capacity) as
select district_name, sum(capacity)
from water_schemes
group by district_name
with total_avg(capacity) as
select avg(capacity)
from total
select name
```

SOLUTIONS

Sol 1. (26)
Result of inner query

Dept Name	Number
AA	4
AB	3
AC	3
AD	2
AE	1

Sol 2. (7)
The ouput of the query is ta.player
Klose
Ronaldo
G Muller
Fontaine
Pele
Klismann
Koesis

Sol 3. (2)

Total	
Name	**Capacity**
Ajmer	20
Bikaner	40
Churu	30
Dungargargh	10

Total_avg
Capacity
25

The result of the query is: name
Bikaner, Churu

Sol 4. (b)
Because in SQL SELECT command retains duplicates by default. In order to eliminate those duplicates we have to write DISTINCT Keyword i.e SELECT DISTINCT

Sol 5. (a)

Sol 6. (a)

Sol 7. (c)

Sol 8. (a)
Optimized version is $\pi_{A_1}\left(\left(\sigma_{F_1 \wedge F_2}\right)(r)\right)$

Sol 9. (d)

Sol 10. (d)
So, an employee whose
ALL customers gives him GOOD rating is chose. All such employees are chose.

Sol 11. (b)

Sol 12. (a)
All the four queries will select the Sname of students with no duplication having Roll number same in both R and S Table and course number as 107 and percentage greater than 90.

Sol 13. (c)
Q and R are True about SQL Query.

Sol 14. (a)
$(A \cup B) \bowtie_{A.id>40 \vee c.id<15} C$

$(A \cup B) \bowtie C \Rightarrow$

A.id	Name	Age	C.id	Name	Age
12			10		
12			99		
15			10		
15			99		
25			10		
25			99		
98			10		
98			99		
99			10		
99			99		

We have to deal with id only. So, tuples with A.id > 40 or C.id < 15 will be selected.