CST 204	Database Management Systems	CATEGORY	L	T	P	CREDIT	YEAR OF INTRODUCTION
		PCC	3	1	0	4	2019

Preamble: This course provides a clear understanding of fundamental principles of Database Management Systems (DBMS) with special focus on relational databases to the learners. The topics covered in this course are basic concepts of DBMS, Entity Relationship (ER) model, Relational Database principles, Relational Algebra, Structured Query Language (SQL), Physical Data Organization, Normalization and Transaction Processing Concepts. The course also gives a glimpse of the alternative data management model, NoSQL. This course helps the learners to manage data efficiently by identifying suitable structures to maintain data assets of organizations and to develop applications that utilize database technologies.

Prerequisite: Topics covered under the course Data Structures (CST 201), Exposure to a High Level Language like C/python.

Course Outcomes: After the completion of the course the student will be able to

CO1	Summarize and exemplify fundamental nature and characteristics of database systems **(Cognitive Knowledge Level: Understand)**
CO2	Model real word scenarios given as informal descriptions, using Entity Relationship diagrams. **(Cognitive Knowledge Level: Apply**)
CO3	Model and design solutions for efficiently representing and querying data using relational model (**Cognitive Knowledge Level: Analyze**)
CO4	Demonstrate the features of indexing and hashing in database applications **(Cognitive Knowledge Level: Apply**)
CO5	Discuss and compare the aspects of Concurrency Control and Recovery in Database systems **(Cognitive Knowledge Level: Apply**)
CO6	Explain various types of NoSQL databases (**Cognitive Knowledge Level: Understand**)

Mapping of course outcomes with program outcomes

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	✓	✓	✓									✓
CO2	✓	✓	✓	✓								✓
CO3	✓	✓	✓	✓								✓
CO4	✓	✓	✓							✓		✓
CO5	✓	✓	✓							✓		✓
CO6	✓	✓	✓		✓					✓		✓

Abstract POs defined by National Board of Accreditation			
PO#	**Broad PO**	**PO#**	**Broad PO**
PO1	Engineering Knowledge	**PO7**	Environment and Sustainability
PO2	Problem Analysis	**PO8**	Ethics
PO3	Design/Development of solutions	**PO9**	Individual and team work
PO4	Conduct investigations of complex problems	**PO10**	Communication
PO5	Modern tool usage	**PO11**	Project Management and Finance
PO6	The Engineer and Society	**PO12**	Life long learning

Assessment Pattern

Bloom's Category	Continuous Assessment Tests		End Semester Examination Marks (%)
	Test1 (%)	Test2 (%)	
Remember	30	30	30
Understand	40	40	40
Apply	30	30	30

Analyze			
Evaluate			
Create			

Mark Distribution

Total Marks	CIE Marks	ESE Marks	ESE Duration
150	50	100	3 hours

Continuous Internal Evaluation Pattern:

Attendance : 10 marks

Continuous Assessment Tests : 25 marks

Continuous Assessment Assignment : 15 marks

Internal Examination Pattern:

Each of the two internal examinations has to be conducted out of 50 marks

First Internal Examination shall be preferably conducted after completing the first half of the syllabus and the Second Internal Examination shall be preferably conducted after completing remaining part of the syllabus.

There will be two parts: Part A and Part B. Part A contains 5 questions (preferably, 2 questions each from the completed modules and 1 question from the partly covered module), having 3 marks for each question adding up to 15 marks for part A. Students should answer all questions from Part A. Part B contains 7 questions (preferably, 3 questions each from the completed modules and 1 question from the partly covered module), each with 7 marks. Out of the 7 questions in Part B, a student should answer any 5.

End Semester Examination Pattern:

There will be two parts; Part A and Part B. Part A contains 10 questions with 2 questions from each module, having 3 marks for each question. Students should answer all questions. Part B contains 2 questions from each module of which a student should answer any one. Each question can have maximum 2 sub-divisions and carries 14 marks.

Syllabus

Module 1: Introduction & Entity Relationship (ER) Model

Concept & Overview of Database Management Systems (DBMS) - Characteristics of Database system, Database Users, structured, semi-structured and unstructured data. Data Models and Schema - Three Schema architecture. Database Languages, Database architectures and classification.

ER model **-** Basic concepts, entity set & attributes, notations, Relationships and constraints, cardinality, participation, notations, weak entities, relationships of degree 3.

Module 2: Relational Model

Structure of Relational Databases - Integrity Constraints, Synthesizing ER diagram to relational schema

Introduction to Relational Algebra - select, project, cartesian product operations, join - Equi-join, natural join. query examples, introduction to Structured Query Language (SQL), Data Definition Language (DDL), Table definitions and operations – CREATE, DROP, ALTER, INSERT, DELETE, UPDATE.

Module 3: SQL DML (Data Manipulation Language), Physical Data Organization

SQL DML (Data Manipulation Language) - SQL queries on single and multiple tables, Nested queries (correlated and non-correlated), Aggregation and grouping, Views, assertions, Triggers, SQL data types.

Physical Data Organization - Review of terms: physical and logical records, blocking factor, pinned and unpinned organization. Heap files, Indexing, Singe level indices, numerical examples, Multi-level-indices, numerical examples, B-Trees & B+-Trees (structure only, algorithms not required), Extendible Hashing, Indexing on multiple keys – grid files.

Module 4: Normalization

Different anomalies in designing a database, The idea of normalization, Functional dependency, Armstrong's Axioms (proofs not required), Closures and their computation, Equivalence of Functional Dependencies (FD), Minimal Cover (proofs not required). First Normal Form (1NF), Second Normal Form (2NF), Third Normal Form (3NF), Boyce Codd Normal Form (BCNF), Lossless join and dependency preserving decomposition, Algorithms for checking Lossless Join (LJ) and Dependency Preserving (DP) properties.

Module 5: Transactions, Concurrency and Recovery, Recent Topics

Transaction Processing Concepts - overview of concurrency control, Transaction Model, Significance of concurrency Control & Recovery, Transaction States, System Log, Desirable Properties of transactions.

Serial schedules, Concurrent and Serializable Schedules, Conflict equivalence and conflict serializability, Recoverable and cascade-less schedules, Locking, Two-phase locking and its variations. Log-based recovery, Deferred database modification, check-pointing.

Introduction to NoSQL Databases, Main characteristics of Key-value DB (examples from: Redis), Document DB (examples from: MongoDB)

Main characteristics of Column - Family DB (examples from: Cassandra) and Graph DB (examples from : ArangoDB)

Text Books

1. Elmasri R. and S. Navathe, Database Systems: Models, Languages, Design and Application Programming, Pearson Education, 2013.

2. Sliberschatz A., H. F. Korth and S. Sudarshan, Database System Concepts, 6/e, McGraw Hill, 2011.

Reference Books:

1. Adam Fowler, NoSQL for Dummies, John Wiley & Sons, 2015
2. NoSQL Data Models: Trends and Challenges (Computer Engineering: Databases and Big Data), Wiley, 2018
3. Web Resource: https://www.w3resource.com/redis/
4. web Resource: https://www.w3schools.in/category/mongodb/
5. Web Resource: https://www.tutorialspoint.com/cassandra/cassandra_introduction.htm
6. Web Resource : https://www.tutorialspoint.com/arangodb/index.htm

Sample Course Level Assessment Questions

Course Outcome1 (CO1):

1. List out any three salient features of database systems, which distinguish it from a file system.
2. Give one example each for logical and physical data independence.

Course Outcome 2(CO2):

1. What facts about the relationships between entities EMPLOYEE and PROJECT are conveyed by the following ER diagram?

1. Design an ER diagram for the following scenario:
   There is a set of teams, each team has an ID (unique identifier), name, main stadium, and to which city this team belongs. Each team has many players, and each player belongs to one team. Each player has a number (unique identifier), name, DoB, start year, and shirt number that he uses. Teams play matches, in each match there is a host team and a guest team.

Course Outcome 3(CO3):

1. For the SQL query, SELECT *A, B FROM R WHERE B='apple' AND C = 'orange'* on the table R(A, B, C, D), where A is a key, write any three equivalent relational algebra expressions.
2. Given the FDs $P \rightarrow Q$, $P \rightarrow R$, $QR \rightarrow S$, $Q \rightarrow T$, $QR \rightarrow U$, $PR \rightarrow U$, write the sequence of *Armstrong's Axioms* needed to arrive at the following FDs: (a) $P \rightarrow T$ (b) $PR \rightarrow S$ (c) $QR \rightarrow SU$
3. Consider a relation PLAYER (PLAYER-NO, PLAYER-NAME, PLAYER-POSN, TEAM, TEAM-COLOR, COACH-NO, COACH-NAME, TEAM-CAPTAIN). Assume that PLAYER-NO is the *only* key of the relation and that the following dependencies hold:

   TEAM→{TEAM-COLOR, COACH-NO, TEAM-CAPTAIN}

   COACH-NO→COACH-NAME.

   i. Is the relation in 2NF? If not, decompose to 2NF.
   ii. Is the relation in 3NF? If not, decompose to 3NF.

4. In the following tables foreign keys have the same name as primary keys except DIRECTED-BY, which refers to the primary key ARTIST-ID. Consider only *single-director* movies.
MOVIES(<u>MOVIE-ID</u>, MNAME, GENRE, LENGTH, DIRECTED-BY)
ARTIST(<u>ARTIST-ID</u>, ANAME)
ACTING(<u>ARTIST-ID, MOVIE-ID</u>)
Write SQL expressions for the following queries:
    (a) *Name(s) and director name(s) of movie(s) acted by 'Jenny'.*
    (b) *Names of actors who have <u>never</u> acted with 'Rony'*
    (c) *Count of movies genre-wise.*
    (d) *Name(s) of movies with maximum length.*

Course Outcome 4(CO4):

1. Consider an EMPLOYEE file with 10000 records where each record is of size 80 bytes. The file is sorted on employee number (15 bytes long), which is the primary key. Assuming un-spanned organization, block size of 512 bytes and block pointer size of 5 bytes. Compute the number of block accesses needed for retrieving an employee record based on employee number if (i) No index is used (ii) Multi-level primary index is used.

Course Outcome 5(CO5):

1. Determine if the following schedule is *recoverable*. Is the schedule *cascade-less*? Justify your answer. *r1(X), r2(Z), r1(Z), r3(X), r3(Y), w1(X), c1, w3(Y), c3, r2(Y), w2(Z), w2(Y), c2*. (*Note: ri(X)/wi(X)* means transaction *Ti* issues read/write on item X; *ci* means transaction *Ti* commits.)
2. Two-phase locking protocol ensures serializability. Justify.

Course Outcome 6(CO6):

1. List out any three salient features of NoSQL databases. Give example of a document in MongoDB.

Model Question paper

QPCODE

Reg No:__________________

Name:___________________

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

FOURTH SEMESTER B.TECH DEGREE EXAMINATION, MONTH & YEAR

Course Code: CST 204

Course Name: **Database Management Systems**

Max.Marks:100 **Duration: 3 Hours**

PART A

Answer all Questions. Each question carries 3 Marks

1 List out any three salient features of a database systems.

2 When is multi-valued composite attribute used in ER modelling?

3 For the SQL query, SELECT *A, B FROM R WHERE B='apple' AND C = 'orange'* on the table R(A, B, C, D), where A is a key, write any two equivalent relational algebra expressions.

4 Outline the concept of *theta*-join.

5 How is the purpose of *where* clause is different from that of having clause?

6 What is the use of a trigger?

7 When do you say that a relation is not in 1NF?

8 Given the FDs $P \rightarrow Q$, $P \rightarrow R$, $QR \rightarrow S$, $Q \rightarrow T$, $QR \rightarrow U$, $PR \rightarrow U$, write the sequence of Armstrong's Axioms needed to arrive at a. $P \rightarrow T$ b. $PR \rightarrow S$

9 What is meant by the lost update problem?

10 What is meant by check pointing?

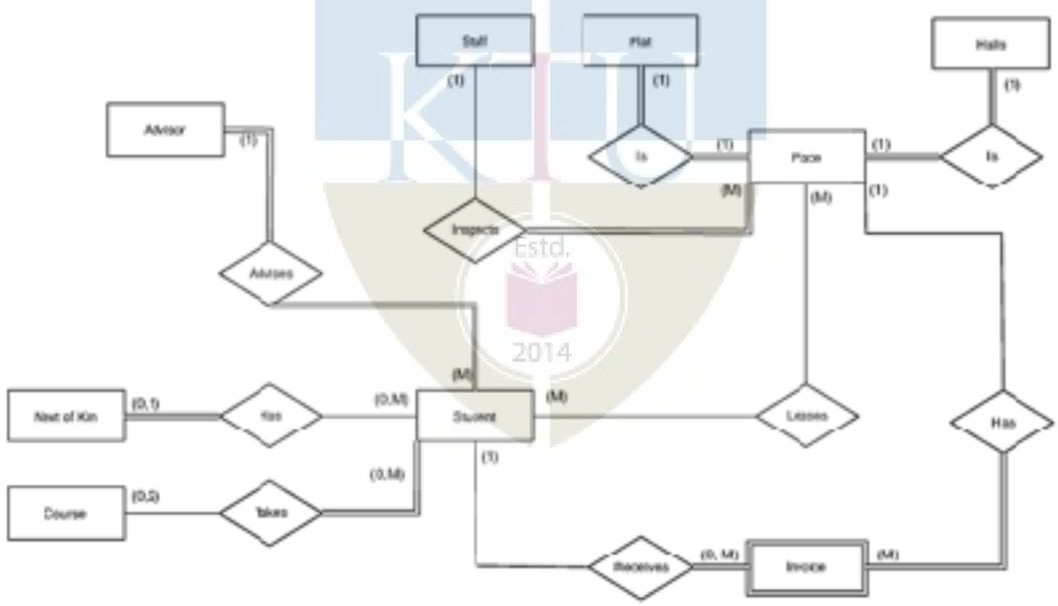
PART B

Answer any one Question from each module. Each question carries 14 Marks

11 a. Design an ER diagram for the following scenario: There is a set of teams, each team has an ID (unique identifier), name, main stadium, and to which city this team belongs. Each team has many players, and each player belongs to one team. Each player has a number (unique identifier), name, DoB, start year, and shirt number that he uses. Teams play matches, in each match there is a host team and a guest team. The match takes place in the stadium of the host team. For each match we need to keep track of the following: The date on which the game is played The final result of the match. The players participated in the match. For each player, how many goals he scored, whether or not he took yellow card, and whether or not he took red card. During the match, one player may substitute another player. We want to capture this substitution and the time at which it took place. Each match has exactly three referees. For each referee we have an ID (unique identifier), name, DoB, years of experience. One referee is the main referee and the other two are assistant referee. (14)

OR

12 a. Interpret the the following ER diagram. (8)



b. Distinguish between physical data independence and logical data independence with suitable examples. (6)

13 **EMPLOYEE(ENO, NAME, ADDRESS, DOB, AGE, GENDER, SALARY, DNUM, SUPERENO)** (14)
DEPARTMENT(DNO, DNAME, DLOCATION, DPHONE, MGRENO)
PROJECT(PNO, PNAME, PLOCATION, PCOST, CDNO)

DNUM is a foreign key that identifies the department to which an employee belongs. MGRENO is a foreign key identifying the employee who manages the department. CDNO is a foreign key identifying the department that controls the project. SUPERENO is a foreign key identifying the supervisor of each employee.

Write relational algebra expressions for the following queries:-

(a) Names of female employees whose salary is more than 20000.
(b) Salaries of employee from 'Accounts' department
(c) Names of employees along with his/her superviser's name
(d) For each employee return name of the employee along with his department name and the names of projects in which he/she works
(e) Names of employees working in all the departments

OR

14 a.Write SQL DDL statements for the the following (Assume suitable domain types): (10)

i. Create the tables STUDENT(ROLLNO, NAME, CLASS, SEM, ADVISER), FACULTY(FID, NAME, SALARY, DEPT). Assume that ADVISER is a foreign key referring FACUTY table.

ii. Delete department with name 'CS' and all employees of the department.

iii. Increment salary of every faculty by 10%.

b.Illustrate foreign key constraint with a typical example. (4)

15 For the relation schema below, give an expression in SQL for each of the queries that follows: (14)

employee(<u>employee-name</u>, street, city)
works(<u>employee-name, company-name</u>, salary)
company(<u>company-name</u>, city)
manages(<u>employee-name, manager-name</u>)

a) Find the names, street address, and cities of residence for all employees who work for the Company 'RIL Inc.' and earn more than $10,000.
b) Find the names of all employees who live in the same cities as the companies for which they work.
c) Find the names of all employees who do not work for 'KYS Inc.'. Assume that all people work for exactly one company.
d) Find the names of all employees who earn more than every employee of 'SB Corporation'. Assume that all people work for at most one company.
e) List out number of employees company-wise in the decreasing order of number of employees.

OR

16 a. Consider an EMPLOYEE file with 10000 records where each record is of size 80 bytes. The file is sorted on employee number (15 bytes long), which is the primary key. Assuming un-spanned organization and block size of 512 bytes compute the number of block accesses needed for selecting records based on employee number if, (9)

i. No index is used
ii. Single level primary index is used
iii. Multi-level primary index is used

Assume a block pointer size of 6 bytes.

b. Illustrate correlated and non-correlated nested queries with real examples. (5)

17 a. Illstrate3NF and BCNF with suitable real examples. (6)

b. Given a relation R(A1,A2,A3,A4,A5) with functional dependencies A1→A2A4 and A4→A5, check if the decomposition R1(A1,A2,A3), R2(A1,A4), R3(A2,A4,A5) is lossless. (8)

OR

18 a. Consider the un-normalized relation R(A, B, C, D, E, F, G) with the FDs A→B , AC→G, AD→EF, EF→G, CDE→AB. Trace the normalization process to reach 3NF relations. (7)

b. Illustrate Lossless Join Decomposition and Dependency Preserving Decomposition with typical examples. (7)

19 a. Discuss the four ACID properties and their importance. (7)

b. Determine if the following schedule is conflict serializable. Is the schedule recoverable? Is the schedule cascade-less? Justify your answers. (7)

r1(X), r2(Z), r1(Z), r3(X), r3(Y), w1(X), c1, w3(Y), c3, r2(Y), w2(Z), w2(Y), c2

(Note: ri(X)/wi(X) means transaction Ti issues read/write on item X; ci means transaction Ti commits.)

OR

20 a. Discuss the main characteristics of Key-value DB and Graph DB. (7)

b. Illustrate two-phase locking with a schedule containing three transactions. Argue that 2PL ensures serializability. Also argue that 2Pl can lead to deadlock. (7)

Teaching Plan

	Course Name	Hours (48)
	Module 1: Introduction & ER Model	**8**
1.1	Concept & Overview of DBMS, Characteristics of DB system, Database Users.	1
1.2	Structured, semi-structured and unstructured data. Data Models and Schema	1
1.3	Three-Schema-architecture. Database Languages	1
1.4	Database architectures and classification	1
1.5	ER model: basic concepts, entity set & attributes, notations	1
1.6	Relationships and constraints – cardinality, participation, notations	1
1.7	Weak entities, relationships of degree 3	1
1.8	ER diagram – exercises	1
	Module 2: Relational Model	**7**
2.1	Structure of relational Databases, Integrity Constraints	1
2.2	Synthesizing ER diagram to relational schema, Introduction to relational algebra.	1
2.3	Relational algebra: select, project, Cartesian product operations	1
2.4	Relational Algebra: join - Equi-join, Natural join	1
2.5	Query examples	1
2.6	Introduction to SQL, important data types	1
2.7	DDL, Table definitions and operations – CREATE, DROP, ALTER, INSERT, DELETE, UPDATE	1
	Module 3: SQL DML, Physical Data Organization	**11**
3.1	SQL DML, SQL queries on single and multiple tables	1
3.2	Nested queries (correlated and non-correlated)	1
3.3	Aggregation and grouping	1

	Course Name	Hours (48)
3.4	Views, assertions (with examples)	1
3.5	Triggers (with examples), SQL data types	1
3.6	Review of terms: physical and logical records, blocking factor, pinned and unpinned organization. Heap files, Indexing	1
3.7	Singe level indices, numerical examples	1
3.8	Multi-level-indices, numerical examples	1
3.9	B-Trees and B+Trees (structure only, algorithms not required)	1
3.10	Extendible Hashing	1
3.11	Indexing on multiple keys – grid files	1
	Module 4: Normalization	**8**
4.1	Different anomalies in designing a database, The idea of normalization	1
4.2	Functional dependency, Armstrong's Axioms (proofs not required)	1
4.3	Closures and their computation, Equivalence of FDs, minimal Cover (proofs not required).	1
4.4	1NF, 2NF	1
4.5	3NF, BCNF	1
4.6	Lossless join and dependency preserving decomposition	1
4.7	Algorithms for checking Lossless Join and Dependency preserving properties (Lecture 1)	1
4.8	Algorithms for checking Lossless Join and Dependency preserving properties (Lecture 2)	1
	Module 5: Transactions, Concurrency and Recovery, Recent Topics	**14**
5.1	Transaction Processing Concepts: Transaction Model	1
5.2	Overview of concurrency control, Significance of concurrency Control & Recovery	1
5.3	Transaction States, System Log	1

	Course Name	Hours (48)
5.4	Desirable Properties of transactions, Serial schedules	1
5.5	Concurrent and Serializable Schedules	1
5.6	Conflict equivalence and conflict serializability	1
5.7	Recoverable and cascade-less schedules	1
5.8	Locking, Two-phase locking, strict 2PL.	1
5.9	Log-based recovery	1
5.10	Deferred database modification (serial schedule), example	1
5.11	Deferred database modification (concurrent schedule) example, check-pointing	1
5.12	Introduction to NoSQL Databases	1
5.13	Main characteristics of Key-value DB (examples from: Redis), Document DB (examples from: MongoDB) [detailed study not expected]	1
5.14	Main characteristics of Column-Family DB (examples from: Cassandra) and Graph DB (examples from : ArangoDB) [detailed study not expected]	1