Part 1: Multiple Choice (120 points - 3 points per question)

(C) 1. One application of data warehouses is:
   (A) shipping of information. (B) order processing. (C) decision support. (D) file updating.

(D) 2. Which is true?
   (A) A database is an organized collection of physically related data. (B) Data after processing is knowledge.
   (C) Database schema changes every time the database is updated. (D) None of the above

(B) 3. All of the following are properties of metadata except:
   (A) data definitions. (B) processing logic. (C) rules or constraints. (D) data structures.

(A) 4. Which of the following is a characteristic of the database approach?
   (A) It stresses the integration of data. (B) It duplicates data.
   (C) It creates physical relationships. (D) It reduces storage.

(C) 5. With the database approach, data descriptions are stored in a central location known as a(n):
   (A) index. (B) data warehouse. (C) repository. (D) data server.

(B) 6. Which is an entity in a bookstore? (A) open hours (B) staff (C) address (D) none of the above

(B) 7. Which of the following is NOT a component of an information systems architecture?
   (A) Data (B) Hardware (C) Networks (D) Reasons for events and rules

(B) 8. Which is a job of a DBA?
   (A) Defining the database constraints (B) Authorizing access to the database
   (C) Defining the database transactions (D) None of the above

(C) 9. The SDLC phase in which functional data specifications and processing rules are created is the _____ phase.
   (A) planning (B) design (C) analysis (D) implementation

(A) 10. Software that provides automated support for some phases of the SDLC is called:
   (A) CASE. (B) CAD. (C) RAD. (D) MST.

(A) 11. Which schemas is in the three-schema approach? (A) internal (B) logical (C) functional (D) dissecting

(C) 12. Which is in the DBMS-dependent design process?
   (A) Requirement analysis (B) Conceptual design (C) Transaction implementation (D) None of the above

(A) 13. Which belongs to the client tier? (A) Web browser (B) Web server (C) database system (D) None of the above

(A) 14. Which is usually to represent a relationship in an ER diagram? (A) verb (B) noun (C) adjective (D) proposition

(A) 15. In the following diagram, which answer is true?

(A) Each employee can supervise one to many employees. (B) Each employee can manage many departments.
   (C) Each employee works in more than one department. (D) none of the above

(A) 16. The following figure shows an example of:
   (A) a composite attribute. (B) a relational attribute. (C) a derived attribute. (D) a multivalued attribute.

(A) 17. Which specifies the number of instances of one entity that can be associated with each instance of another entity?
   (A) cardinality (B) degree (C) identification (D) participation

(B) 18. Which type of relationships between a doctor and a patient?
   (A) one-to-many (B) many-to-many (C) many-to-one (D) one-to-one

(A) 19. Which of the following is true?
   (A) The EER is a type of conceptual data models. (B) C. J. Date first proposed the relation model.
   (C) An entity can exist in the database merely by being a member of a subclass. (D) None of the above

(B) 20. The process of defining one or more subtypes of a supertype and forming relationships is called:
   (A) specialization. (B) generalization. (C) creating discord. (D) selecting classes.

(C) 21. Which rule states that an entity instance cannot simultaneously be a member of two (or more) subtypes
   (A) partial specialization (B) overlap (C) disjoint (D) total specialization
22. Which of the following is false?
   (A) A candidate key can uniquely identify a row. (B) Key constraint means the primary key cannot have null values.
   (C) A relation can have multiple candidate keys. (D) An enterprise key whose value is unique across all relations.

23. In the SQL language, which statement can be used to implement business rules in databases?
   (A) create operation (B) select constraint (C) update business rule (D) create assertion

24. In the following EER diagram, which is true?

   - A student must be a graduate student or an undergraduate student.
   - A student must be a graduate student, an undergraduate, or a special student.
   - A student must be at least a special student.
   - None of the above

25. The relational data model consists of which components?
   (A) Data structure (B) Data manipulation (C) Data integrity (D) All of the above

26. Which constraint may the delete operation violate?
   (A) Referential constraint (B) Entity constraint (C) Key integrity (D) None of the above

27. Which of the following finds those groups meeting stated conditions?
   (A) group by (B) having (C) using by (D) none of the above

28. Which is an approach to map a n-ary relationship type in the ER diagram to the relational schema?
   (A) Merged relation (B) A relationship relation and n foreign keys (C) A relation set of simple component attributes
   (D) None of the above

29. To eliminate duplicate rows in a query, which can be used?
   (A) unique (B) index (C) distinct (D) none of the above

30. Which integrity constraints can trigger a sequence of operations?
   (A) restrict (B) set default (C) set null (D) cascade

31. Which is a join condition in the following SQL commands?

   ```sql
   select name from employee, department
   where department.name = 'Research' and employee.department_no = department.department_no;
   ```

   (A) department.name = 'Research' (B) employee.department_no = department.department_no
   (C) select name from employee, department (D) None of the above

32. In MySQL which is used to execute a SQL script? (A) use (B) \e (C) \. (D) none of the above

33. Which is a benefit of denormalization?
   (A) performance improvement (B) higher security (C) less storage (D) none of the above

34. A rule that database users should obey is called a: (A) regulation. (B) principle. (C) privilege. (D) constraint.

35. A join that is based upon equality between values in two common columns with the same name and where one
   duplicate column has been removed is called a(n):

   (A) equi-join. (B) natural join. (C) multivariate join. (D) inner join.

36. Which is a subset of the database that is presented to one or more users?
   (A) correlated subquery (B) derived table (C) view table (D) None of the above

37. What results would the following SQL statement produce?

   ```sql
   select owner, table_name from dba_tables where table_name = 'customer';
   ```

   (A) A listing of all customers in the customer table (B) A listing of the owner of the customer table
   (C) A listing of the owner of the customer table as well as customers (D) An error message

38. Which provides a standard software API for using DBMS? (A) ODBC (B) CASE (C) OLAP (D) none of the above

39. The UNION clause is used to:

   (A) find all rows that do not match in two tables.
   (B) join two tables together to form one table.
   (C) combine the output from multiple queries into a single result table.
   (D) None of the above.

40. Web services:

   (A) are a set of business processes available on the Web. (B) are a set of standards based upon HTML.
   (C) are a set of standard protocols for automatic communication between software over the Web. (D) none of the above.

Part 2: Questions and Answers (126 points)
1. (21 points) Briefly explain these terminologies. If they are acronyms, also write what they stand for.
   (a) data model (b) entity instance (c) functional dependency (d) identifying relationship (e) data mining (f) XML (g) XHTML
   (a) A set of concepts to describe the structure of a database, the operations for manipulating these structures, and certain
       constraints that the database should obey.
   (b) An entity (instance) is a specific object or thing in the mini-world that are represented in the database.
   (c) Functional dependency specifies that the value of one attribute (the determinant) determines the value of another
       attribute in the same table.
   (d) An identifying relationship is the relationship between a weak entity type and its owner.
   (e) The data mining can be defined in either one as shown in below:
       • The discovery of new information in terms of patterns or rules from vast amounts of data.
       • The process of finding interesting structure in data.
       • The process of employing one or more computer learning techniques to automatically analyze and extract knowl-
         edge from data.
   (f) EXtensible Markup Language (XML) is a language used to specify the data content.
   (g) The Extensible Hypertext Markup Language is a markup language to display documents on the Web as an application
       of the Extensible Markup Language (XML).

2. (12 points) Briefly explain the four constraints on specialization and generalization and give an example for each case.
   • The disjointness constraint specifies that the subclasses of the specialization must be disjoint.
     A graduate student cannot be a undergraduate at the same time.
   • The overlapping specialization specifies that the subclasses of the specialization can be overlapping.
     A manager can be an engineer at the same time.
   • The total specialization specifies that every entity in the superclass must be a member of some subclass.
     Human beings can be classified into males and females.
   • The partial specialization specifies that an entity in the superclass is allowed not to belong to any of the subclasses.
     A student need not to be an undergraduate or a graduate student.

3. (8 points) Explain the key constraint, domain constraint, entity integrity constraint, and referential constraint.
   • The key constraint means there is no duplicate key in any relation.
   • The domain constraint indicates every value in a tuple must be from the domain of its attribute.
   • The entity integrity indicates the values of primary key attributes in a relation cannot be null.
   • The referential integrity constraints indicate any attribute of a foreign key in a table can contain only either values
     from the corresponding parent table’s primary key or the null value.

4. (a) (4 points) What does SQL stand for? Explain it.
   (b) (6 points) Based on the functions how can SQL be classified into three categories?
   (a) Structured Query Language (SQL) is a standard language used to retrieve, update and delete data from relational
       database management systems (DBMS).
   (b) Data Definition Language (DDL) is used to define databases.
       Data Manipulation Language (DML) is used to manipulate databases.
       Data Control Language (DCL) is used to control databases.

5. (a) (6 points) Describe three types of anomalies that can arise in a table.
   (b) (3 points) Explain normalization for a relational database.
   (c) (6 points) Explain the 1NF, 2NF, and 3NF.
   (a) • An insertion anomaly is an anomaly in which adding new rows forces user to create duplicate data.
       • A deletion anomaly is an anomaly in which deleting rows may cause a loss of data that would be needed for other
         future rows.
       • A modification anomaly is an anomaly in which changing data in a row forces changes to other rows because of
         duplication.
   (b) The process of structuring relations by decomposing their attributes into smaller relations.
   (c) • 1NF is the relation that has no composite attributes, multivalued attributes, and nested relations.
       ii. In 2NF, every non-prime attribute is fully functionally dependent on the primary key in the relation.
       Other possible answers:
       There are no two keys in the relation.
       There is no partial functional dependency in the relation.
iii. Third normal form (3NF) is a normal form in which there is no transitive functional dependency in the relation.

6. (24 points) Consider the following database for a library:

<table>
<thead>
<tr>
<th>student table</th>
<th>checkout table</th>
</tr>
</thead>
<tbody>
<tr>
<td>student_id</td>
<td>student_id</td>
</tr>
<tr>
<td>name</td>
<td>isbn</td>
</tr>
<tr>
<td>department</td>
<td>checkout_date</td>
</tr>
<tr>
<td>B96022</td>
<td>B96022</td>
</tr>
<tr>
<td>Taylor Swift</td>
<td>0130428981</td>
</tr>
<tr>
<td>Marketing</td>
<td>2009-6-6</td>
</tr>
<tr>
<td>B96145</td>
<td>B96145</td>
</tr>
<tr>
<td>Lily Allen</td>
<td>0316172324</td>
</tr>
<tr>
<td>Music</td>
<td>2009-6-16</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>book table</th>
</tr>
</thead>
<tbody>
<tr>
<td>isbn</td>
</tr>
<tr>
<td>0130428981</td>
</tr>
<tr>
<td>0316172324</td>
</tr>
</tbody>
</table>

where primary keys are underlined. student_id and isbn in the checkout table are foreign keys referencing to the student and book table respectively.

(a) If the following operations are taken, check if domain constraints, key constraints, entity integrity, or referential integrity is violated. If there is any violation, explain it.

i. (2 points) Insert ('B96022', 'Lady Gaga', 'History') into the student table.
ii. (2 points) Change the student_id of 'B96145' in the checkout table from 'B96145' to 'B96022'.
iii. (2 points) Remove the row ('0316172324', 'Blink', 'M. Gladwell', 2004) from the book table.
iv. (2 points) Change the title in the book table from 'Advanced Database' to 'Advanced Database Systems'.

(b) Use SQL to answer the following questions.

i. (3 points) Create the student table with the required constraints.
ii. (3 points) Add a constraint of \(0 \leq \text{duration} \leq 365\) in the checkout table.
iii. (2 points) Insert ('B96022', 'Lady Gaga', 'History') into the student table.
iv. (3 points) For each student, list the student name and the total number of books borrowed by the student.
v. (2 points) Change the department of Taylor Swift from 'Marketing' to 'English'.
vi. (3 points) Remove all books borrowed by Taylor Swift.

(a) i. It violates the key constraint because the student_id 'B96022' already existed.
ii. It violates no constraint.
iii. It violates the referential integrity because the foreign key, isbn '0316172324' in the checkout table will have no primary key to reference to in the department table.
iv. It violates no constraint.

(b) i. create table student (student_id char(6) primary key not null, name varchar(30), department varchar(30));
ii. alter table checkout add constraint duration_constraint check (0 <= duration and duration <= 40);
iii. insert into student values ('B96022', 'Lady Gaga', 'History')
iv. select name, count(*) from student, checkout where student.student_id = checkout.student_id by stduent_id;
v. update student set department = 'English' where name = 'Taylor Swift'
vi. delete from book where isbn in (select isbn from student, checkout where student.student_id = checkout.student_id and name = 'Taylor Swift')

7. (20 points) Consider a HollyWood Enterprise that requires modeling information about the different type of people involved in the movie production.

- Each person should have person ID, name, phone, gender, and address.
- There are two main groups of persons: Movie professionals and celebrity. Each movie professionals work on some company.
- A movie professionals can be either a director or a agent. Each director has her or his popularity and can direct a number of movies. Each agent has the agent fee.
- A celebrity can be a movie star, a model, or both. Each movie star has her or his movie style and play in some movies. Each model has her or his preferences.
- Each movie has the information about the title, released date, and language.

(a) (10 points) Draw a EER diagram for the above HollyWood Enterprise.
(b) (10 points) Transform the above EER diagrams to 3NF relations and show the 3NF relation schema.
8. (16 points) Consider the following schema for the enrollment information. Normalize it to 3NF relations.
enrollment(student_id, student_name, course_no, title, credit, section_no, classroom, time, employee_id, instructor_name, office)
where primary keys are underlined.
(a) (8 points) Draw the ER diagram.
(b) (8 points) Show the relation schema.