PART A. Short Answers. (5 points each, 50 points total)

A.1) Consider relation Employee(ID, salary, departmentID), where ID is the key of the relation. Suppose you own this relation, and you would like to give user "Sally" authorization to select records from Employee, but only records for employees whose salary is less than $80,000 and who work in a department with less than 5 people. Give a sequence of SQL commands that achieves this goal.

A.2) Stored Procedures. Database systems have stored procedures. Give three distinct advantages of stored procedures writing one sentence per advantage. (5 points)

A.3) Peer-to-Peer Systems. As we discussed in class, the Gnutella protocol floods each query for a given number of hops throughout the system. In at most two sentences each, describe briefly two approaches how you could improve the search for files in a peer-to-peer system. Your answer will consist of at most four sentences. (5 points)

A.4) Normalization. Give an example of a relation schema R together with three non-trivial FDs that hold over R such that R is in 3NF but not in BCNF. An example **incorrect answer** would be: R(A,B,C) with A→B and B→C. (5 points)

A.5) Querying. Consider the following schema:
Suppliers(sid: integer, sname: string, address: string)
Parts(pid: integer, pname: string, color: string)
Catalog(sid: integer, pid: integer, price: real)

Describe in one concise English sentence what the following SQL query computes. (5 points)

SELECT P.pid, AVG(P.price)
FROM Parts P, Catalog C
WHERE P.pid = C.pid
GROUP BY P.pid
HAVING COUNT(*)>1

A.6) ER-Modeling. Draw an ER-diagram that captures the following two SQL statements. Do not forget to add relevant key constraints and participation constraints. You do not need to write any text for this question. (5 points)

CREATE TABLE Employees (ssn   CHAR(11),
                          Name   CHAR(30),
                          mlot   INTEGER,
                          PRIMARY KEY(ssn))

CREATE TABLE Dept_Mgr (did   INTEGER,
                         dname CHAR(20),
                         ssn   CHAR(11) NOT NULL,
                         PRIMARY KEY (did),
                         FOREIGN KEY (ssn) REFERENCES Employees,

ON DELETE NO ACTION)

A.7) Query Evaluation. Consider a relation R(A,B) with two integer attributes A and B. Assume that the only index on relation R is an unclustered B+-tree on attribute A. R has 1,000,000 pages on disk. The attribute values of attribute A are uniformly distributed between 1 and 1,000. There are 1000 records per page. Consider the following query:

SELECT A, B FROM R WHERE A> 990

Describe or draw the most efficient plan to evaluate this query, and write in at most two sentences why you think this is the most efficient plan. (5 points)

A.8) Transactions. List the ACID properties of a transaction, and describe each property in exactly one concise English sentence. (5 points)

A.9) Query Evaluation. Assume that you are given relations R(A,B,C), S(C,D,E), and T(E,F,G). Assume that each relation has 1,000,000 records, and that each field (A, B, …, G) has 100 distinct attribute values that all appear equally frequently. Assume that you are given the following query:

SELECT T.E, T.G
FROM R, S, T
WHERE R.C=S.C AND S.E = T.E AND T.E = 100 and T.G = > 100

What index(es) would you create to speed up this query, and what is the query plan that you envision with the index(es) that you created? The answer is a query plan and the indexes that are used in the query plan. (5 points)

A.10) Three-tier Architectures. Describe three advantages of a three-tier architecture versus a two-tier architecture in one sentence each. (5 points)

PART B. Queries (15 points)

Consider the following schema (keys are underlined):

Customer(cid, cname, age, gender), Buys(cid, pid), Product(pid, name, type, mfgr, price)

B.1) Write the following query **in relational algebra without using the division operator**: "Find the names of all customers who have purchased all products manufactured by Sears." (5 points)

B.2) Write the following query **in SQL**: "Find the names of all customers who have not purchased the second most expensive product." (5 points)

B.3) Write the following query **in SQL**: "Find the names of all customers who have purchased all products". (5 points)

PART C. Decision Support (10 points)

C.1) Describe in one sentence each three differences between the query workload of an OLTP (Online Transaction Processing) system and a Data Warehouse. (5 points)

C.2) Give an example star schema with one fact table and three dimension tables. The fact table has to have two measures. Each dimension table has to have four attributes. Your answer should be four tables and their respective attributes, where each attribute name is descriptive enough to understand its meaning (the level of detail given in Part B is sufficient). (5 points)

PART D. Data Mining (15 points)

D.2) Assume that you are given the following database. Find all association rules with support >= 50% and confidence >= 60%. (4 points)

TID	CID	Date	Item
111	201	5/1/99	Pen
111	201	5/1/99	Ink
111	201	5/1/99	Milk
111	201	5/1/99	Juice
112	105	6/3/99	Pen
112	105	6/3/99	Ink
112	105	6/3/99	Milk
113	106	6/5/99	Pen
113	106	6/5/99	Milk
114	201	7/1/99	Pen
114	201	7/1/99	Ink
114	201	7/1/99	Juice

PART E. XML Technologies (10 points)

E.1) Write a well-formed and valid XML document that adheres to the DTD below. (5 points)

```
<?xml version='1.0'?>
<!ELEMENT Basket (Cherry+, (Apple | Orange)+) >
        <!ELEMENT Cherry EMPTY>
                <!ATTLIST Cherry flavor CDATA #REQUIRED>
        <!ELEMENT Apple EMPTY>
                <!ATTLIST Apple color CDATA #REQUIRED>
        <!ELEMENT Orange EMPTY>
                <!ATTLIST Orange location 'Florida'>
```

E.2) Write an XSLT transformation that transforms a document that conforms to the DTD above to conform to the DTD below. (5 points)

```
<?xml version='1.0'?>
<!ELEMENT New_Basket (New_Apple, New_Orange)*) >
        <!ELEMENT New_Apple EMPTY>
        <!ELEMENT New_Orange EMPTY>
```