**Module 5**

**Database Design: Normalization**

Module Review Questions

Problems

1. Which of the following is *not* a benefit of the database normalization process?
2. Redundant data is minimized.
3. Data is simplified into only the important attributes.
4. Data is more accurate.
5. Data is more consistent.

2. Which of the following is *not* a type of data modification anomaly?

a. entry

b. update

c. deletion

d. insertion

3. Which of the following best describes functional dependence?

a. It describes the relationship of attributes in an entity.

b. It describes the relationship of tuples in an entity.

c. It describes the relationship of entities as compared to other entities.

d. It describes the relationship of unnormalized data to 1NF.

4. Which of the following is *not* a characteristic of 1NF?

a. Data is stored in a single tabular arrangement.

b. A column may contain only one value.

c. Each column contains an atomic value.

d. An entity is related to another entity through the use of the foreign key field.

5. Which of the following is *not* an atomic value?

a. 15

b. Oklahoma City

c. IBM Corporation

d. Bill Gates

6. Transitive dependencies are removed in which normal form?

a. 1NF

b. 2NF

c. 3NF

d. 4NF

7. Repeating values in attributes are removed in which normal form?

a. 1NF

b. 2NF

c. 3NF

d. 4NF

8. Smaller tables are first removed from a single large list in which normal form?

a. 1NF

b. 2NF

c. 3NF

d. 4NF

9. Which of the following is *not* true of the primary key field of a table?

a. It may not include nulls.

b. It must contain unique data for every record.

c. It must contain numeric data.

d. It may be used in a relationship with another table.

10. What is the purpose of a foreign key field?

a. to link a record to data in another table

b. to shorten the normalization process

c. to uniquely identify each record

d. to prevent unauthorized access to a table

11. Who invented normalization?

a. Thomas John Watson

b. Edgar F. Codd

c. Larry Ellison

d. Grace Hopper

12. Which of the following is least likely to be described as an algorithm?

a. a recipe

b. the Rosetta stone

c. the process of normalization

d. a mathematical formula

Critical Thinking Questions

1. Using your knowledge of the college environment, convert a table with the following notation to an equivalent collection of tables that are in third normal form:

Student (StudentNum, StudentName, (CourseNum, Description, Term, Grade))

Solution:

Students (StudentNum, StudentFirstName, StudentLastName)

Courses (CourseID, Description, Credits)

Enrollments (StudentNum, CourseID, TermID, Grade)

Terms (TermID, Term)

The solution would have at least these four tables and attributes. StudentName should be broken down into two atomic values.

Of particular interest are the foreign key fields in the Enrollments table.

* One student can enroll in multiple classes using the StudentNum field.
* One class can have multiple enrollments using the CourseID field.
* One term can have multiple enrollments using the TermID field.

1. Again, using your knowledge of the college environment, convert the following data about student activities (Chess Club, Marching Band, Ski Club, so forth) into third normal form:

Student (StudentID, StudentName, ActivityNum, ActivityName, Advisor)

Solution:

Students (StudentNum, StudentFirstName, StudentLastName)

Activities (ActivityNum, ActivityName, AdvisorID)

StudentActivities (StudentNum, ActivityNum, JoinDate)

Advisors (AdvisorID, AdvisorFirst, AdvisorLast)

The solution would have at least these four tables and attributes. StudentName and Advisor should be broken down into two atomic values to store the first and last names into two separate fields.

Of particular interest are the foreign key field of AdvisorID in the Activities table to link one advisor to potentially many activities as well as the foreign key fields of StudentNum and ActivityNum in the StudentActivities table to link the record to a given student and activity.

The JoinDate field was added to the StudentActivities table to track the date that the student joined the activity group, but is optional. Other optional date fields or other attributes might be included in a more fully-developed solution.

JC Consulting Case Exercises

The following exercises are based on the JC Consulting database.

Problems

1. What is the primary key field for the Employees table?

a. EmployeeID

b. HireDate

c. Salary

d. LastName

2. Open the Employees table. What field contains redundant data that most likely could be moved into a lookup table?

a. EmployeeID

b. LastName

c. Title

d. Salary

3. Which normal form do these tables adhere to?

a. unnormalized

b. 1NF

c. 2NF

d. 3NF

4. What foreign key fields does the ProjectLineItems table contain?

a. ProjectID and TaskID

b. ProjectLineItemID and ProjectID

c. ProjectLineItemID and TaskID

d. ProjectLineItemID, ProjectID, and TaskID

5. What foreign key fields does the Projects table contain?

a. ClientID and EmployeeID

b. ProjectID and ClientID

c. ProjectID and EmployeeID

d. ProjectID, ClientID, and EmployeeID

6. What foreign key fields does the Clients table contain?

a. ClientID

b. Zip

c. ClientID and Zip

d. The Clients table does not contain any foreign key fields.

7. The Categories table contains one field, Category. Which table will most be used with it to help constrain the values in a particular field?

a. Employees

b. Clients

c. Projects

d. TaskMasterList

8. Open the Relationships window and examine the field lists. What does the key symbol to the left of certain field names represent?

a. the foreign key field for each table

b. the primary key field for each table

c. the composite key field for each table

d. the combination key field for each table

9. Open the TaskMasterList table in Datasheet View. What field contains redundant data that could most likely could be pulled into a lookup table?

a. Per

b. Description

c. TaskID

d. CategoryID or Per

Critical Thinking Questions

1. What is the notation for each of the seven current tables in the JC Consulting database?

Categories (Category)

Clients (ClientID, ClientName, Street, Zip, Government)

Employees (EmployeeID, LastName, FirstName, HireDate, Title, Salary)

ProjectLineItems (ProjectLineItemID, ProjectID, TaskID, TaskDate, Quantity, Factor, ProjectLineItemNotes)

Projects (ProjectID, ProjectStartDate, ClientID, EmployeeID, ProjectNotes)

TaskMasterList (TaskID, Description, CategoryID, Per, Estimate)

Zips (City, Zip)

2. Examine the Employees table. Note that the data in the Title field contains redundant values and therefore is a candidate for a one-field lookup table. Use your Access Query Design View or SQL skills to create a lookup table named Titles with the unique values in the Title field of the Employees table. List the steps you performed to create the lookup table.

Query Design View solution:

1. Start a new query, then add the Employees table to the upper pane of Query Design View.

2. Add the Title field to the first column of the query grid.

3. Click the Totals button in the Show/Hide group of the ribbon to add the Total row to the query grid.

4. Make sure that Group By is the choice in the Total row for the Title field.

5. Click the Make Table button in the Query Type group of the ribbon.

6. Enter Titles as the Table Name value in the Make Table dialog box, then click OK.

7. Click the Run button in the Results group of the ribbon to create the Titles table.

8. Click Yes when prompted that you are about to paste 8 rows into the table.

SQL solution:

1. Start a new query and switch to SQL View.

2. Enter the following SQL statements into the window:

SELECT Title INTO Titles  
FROM Employees  
GROUP BY Title;

3. Click the Run button in the Results group of the ribbon to create the Titles table.

4. Click Yes when prompted that you are about to paste 8 rows into the table.

Pitt Fitness Case Exercises

Problems

1. Suppose Pitt Fitness offered personal fitness training and the Customer table were designed as follows:   
   Customers (CustomerID, LastName, FirstName, StreetAddress, City, State, ZipCode, EmailAddress, PhoneNumber, BirthDate, InstructorID)  
   This design reflects that instructors also work as personal trainers. Which of the following fields is functionally dependent on InstructorID?
   1. LastName
   2. EmailAddress
   3. BirthDate
   4. None of the fields in the table is functionally dependent on InstructorID.
2. Is InstructorLastName a candidate key along with InstructorID in the Instructors table?
   1. Yes, InstructorLastName could be used as a key field because no one instructor has the same last name.
   2. No, InstructorLastName could not be a key field because an instructor might join the team with the same last name as another instructor, such as a sibling.
   3. Yes, InstructorLastName could be used as a key field as long as it was spelled correctly.
   4. Yes, InstructorLastName could be used along with InstructorFirstName as a compound key field.
3. Pitt Fitness is starting to sell its own line of clothing. The following Orders table has been created in the database to record the clothing orders, but the table may not be designed properly. The potential problem may be corrected in which normal form?

**OrderID OrderDate CustomerID ItemID Quantity**

2501 10/1/2021 107 S-61 1

2502 10/1/2021 108 S-59 1

P-25 1

T-19 3

2503 10/1/2021 105 S-45 1

T-91 4

2504 10/1/2021 109 P-24 1

* 1. first normal form
  2. second normal form
  3. third normal form
  4. It is correctly designed.

1. Pitt Fitness modified the Orders table as follows, but it may still not be designed properly. The potential problem may be corrected in which normal form?

**OrderID (key) OrderDate CustomerID ItemID (key) Description Price**

2501 10/1/2021 107 S-61 T-shirt L $25

2502 10/1/2021 108 S-59 T-shirt M $25

2502 10/1/2021 108 P-25 Yoga Pants M $50

2502 10/1/2021 108 T-19 Tank L $18

2503 10/1/2021 105 S-45 T-shirt S $18

2504 10/1/2021 109 P-24 Yoga Pants L $50

* 1. first normal form
  2. second normal form
  3. third normal form
  4. It is correctly designed.

1. Pitt Fitness modified the Reservations table as follows, but it may have an improper design. The potential problem may be corrected in which normal form?

**Reservation Date CustomerID ClassID Instructor LastName FirstName  
(key)**

2100001 1/5/2021 107 39 RS02 Sisto Robert

2100002 1/8/2021 108 24 VP01 Pegues Vicki

2100003 1/10/2021 108 10 RS02 Sisto Robert

2100004 1/10/2021 108 10 NT01 Tahan Neda

2100005 1/12/2021 105 8 NT01 Tahan Neda

2100006 1/13/2021 105 35 VP01 Pegues Vicki

* 1. first normal form
  2. second normal form
  3. third normal form
  4. It is correctly designed.

1. Pitt Fitness asks you to create a query using the following table to find any repeating values of customer and date. What query function in Access should you use when you design this query?

**Reservation Date CustomerID ClassID InstructorID LastName FirstName  
(key)**

2100001 1/5/2021 107 39 RS02 Sisto Robert

2100002 1/8/2021 108 24 VP01 Pegues Vicki

2100003 1/10/2021 108 10 RS02 Sisto Robert

2100004 1/10/2021 108 10 NT01 Tahan Neda

2100005 1/12/2021 105 8 NT01 Tahan Neda

2100006 1/13/2021 105 35 VP01 Pegues Vicki

2100007 1/13/2021 109 33 MS01 Said Memo

* 1. Totals – Group by
  2. Parameters
  3. Append
  4. This cannot be accomplished in an Access query.

1. The Reservations table is organized as follows:   
   Reservations (ReservationID, ClassID, ClassInstructors, ClassPrice, OtherFees, CustomerID)  
   Knowing about normalization, what superior way could you use to organize the data?
   1. Create two tables: Reservations (ReservationID, ClassID, ClassInstructors, CustomerID); ClassPrice (ClassID, ClassPrice, OtherFees)
   2. Create three tables: Reservations (ReservationID, ClassID, ClassInstructors, CustomerID); ClassPrice (ClassID, ClassPrice); ClassOtherFees (ClassID, OtherFees)
   3. Add a new field to the table: Reservations (ReservationID, ClassID, ClassInstructors, PriceID, ClassPrice, OtherFees, CustomerID)
   4. The data is organized in the best possible way.
2. The table ClassInstructors (ClassID, InstructorID) could be considered what type of table?
   1. make table
   2. Codd table
   3. lookup table
   4. anomaly table
3. In the following table, which field does *not* contain an atomic value?

**Reservation Date CustomerID ClassID InstructorID Name  
(key)**

2100001 1/5/2021 107 39 RS02 Robert Sisto

2100002 1/8/2021 108 24 VP01 Vicki Pegues

2100003 1/10/2021 108 10 RS02 Robert Sisto

2100004 1/10/2021 108 10 NT01 Neda Tahan

2100005 1/12/2021 105 8 NT01 Neda Tahan

2100006 1/13/2021 105 35 VP01 Vicki Pegues

2100007 1/13/2021 109 33 MS01 Memo Said

* 1. Reservation
  2. Date
  3. InstructorID
  4. Name

1. What is the notation for the following new table?

**OrderID (key) OrderDate CustomerID**

2501 10/1/2021 107

2502 10/1/2021 108

2502 10/1/2021 108

2502 10/1/2021 108

2503 10/1/2021 105

2503 10/1/2021 105

2504 10/1/2021 109

* 1. Orders (OrderID, OrderDate, CustomerID, ItemID)
  2. Orders (OrderID, OrderDate, CustomerID)
  3. Orders (OrderID, OrderDate, CustomerID)
  4. Orders (OrderID, OrderDate, CustomerID)

Critical Thinking Questions

1. Assume that you designed the field LengthOfTime in the Classes table in the Pitt Fitness database. The LengthOfTime field has data such as 60 minutes, 30 minutes, 1 hour, and 15 minutes. Is this the correct way to record this data? If not, what would be a better method?

Answer: The data in the field should be an atomic value, so you should remove the words minutes and hour and replace the data with numbers only in minutes or hours (or fractions thereof). The module states: An **atomic value** is a piece of data that cannot be meaningfully divided.

1. Originally, Pitt Fitness had all their data in one large Excel spreadsheet. Suppose you were hired to help Pitt Fitness create a database to replace the spreadsheet. Outline the major steps you would perform to create a normalized database from the data in the spreadsheet.

Answer: Step 1 – bring the data into one table in Access and label that raw data. Step 2 – Create a series of make-table queries that split the data into the proper tables to ensure a normalized database, proceeding through 1NF, 2NF, and 3NF.

Sports Physical Therapy Case Exercises

Problems

1. How do you diagram the functional dependence on TherapyCode in the Therapies table?
   1. TherapyCode + Description, UnitOfTime
   2. TherapyCode (Description, UnitOfTime)
   3. TherapyCode 🡪 Description, UnitOfTime
   4. TherapyCode = Description, UnitOfTime
2. What would be a candidate key for the Session table, besides the current key field of SessionNum? The design of the Session table is Session (SessionNum, SessionDate, PatientNum, LengthOfSession, TherapistID, TherapyCode).
   1. SessionDate
   2. PatientNum and TherapistID (compound key)
   3. TherapyCode and PatientNum (compound key)
   4. SessionNum is the only appropriate key field.
3. What foreign key(s) does the Session table contain? Session (SessionNum, SessionDate, PatientNum, LengthOfSession, TherapistID, TherapyCode).
   1. PatientNum, TherapistID, TherapyCode
   2. PatientNum only
   3. PatientNum and TherapistID only
   4. This table does not have any foreign keys.
4. What field is the link between the Therapist and Session tables?
   1. PatientNum
   2. TherapistID
   3. SessionNum
   4. TherapyCode
5. Suppose the Session table is set up as follows. Is this design correct? If not, in which normal form would you correct it?

**Session table**

**SessionNum SessionDate PatientNum LengthOfSession TherapistID Last First TherapyCode**

27 10/10/2021 1011 45 JR085 Risk Jonathan 92507

28 10/11/2021 1016 30 AS648 Shields Anthony 97010

29 10/11/2021 1014 60 SW124 Wilder Steven 97014

30 10/12/2021 1013 30 BM273 McClain Bridgette 97033

31 10/15/2021 1016 90 AS648 Shields Anthony 98960

32 10/16/2021 1018 15 JR085 Risk Jonathan 97035

33 10/17/2021 1017 60 SN852 Nair Saritha 97039

34 10/17/2021 1015 45 BM273 McClain Bridgette 97112

35 10/18/2021 1010 30 SW124 Wilder Steven 97113

36 10/18/2021 1019 75 SN852 Nair Saritha 97116

37 10/19/2021 1020 30 BM273 McClain Bridgette 97124

38 10/19/2021 1021 60 AS648 Shields Anthony 97535

* 1. first normal form
  2. second normal form
  3. third normal form
  4. It is correctly designed; it does not violate any form.

1. Patient Number 1014 is receiving three treatments in one session. Is the following design correct for the Session table?

**Session table**

**SessionNum SessionDate PatientNum LengthOfSession TherapistID TherapyCode**

27 10/10/2021 1011 45 JR085 92507

28 10/11/2021 1016 30 AS648 97010

29 10/11/2021 1014 60 SW124 97014

15 97016

15 97018

30 10/12/2021 1013 30 BM273 97033

31 10/15/2021 1016 90 AS648 98960

32 10/16/2021 1018 15 JR085 97035

33 10/17/2021 1017 60 SN852 97039

34 10/17/2021 1015 45 BM273 97112

35 10/18/2021 1010 30 SW124 97113

36 10/18/2021 1019 75 SN852 97116

37 10/19/2021 1020 30 BM273 97124

38 10/19/2021 1021 60 AS648 97535

1. Yes, this is a good way to record multiple sessions.
2. Yes, but the two additional sessions do not need to be recorded. The time for the original session can just be increased.
3. No, it is not in first normal form (1NF).
4. No, the TherapyCode field is not necessary.
5. Suppose the Therapist table is set up as follows. Why is this table not in 1NF?

**TherapistID LastName Street City State ZipCode**

AS648 Anthony Shields 5222 Eagle Court Palm Rivers TX 72511

BM273 Bridgette McClain 385 West Mill St. Waterville TX 76658

JR085 Jonathan Risk 1010 650 North Palm Rivers TX 72511

SN852 Saritha Nair 25 North Elm St. Livewood TX 72512

SW124 Steven Wilder 7354 Rockville Road San Vista TX 72510

* 1. The table contains repeating groups.
  2. A field does not contain atomic values.
  3. The table includes duplicate records.
  4. The table is in 1NF.

1. When a patient comes into the Sports Physical Therapy, they often have insurance to cover some or all of the cost of the therapy sessions. Assuming that a few patients might use two insurance companies (e.g., a spouse’s insurance might supplement the patient’s insurance), how would you record the insurance information ensuring that the table is normalized?
   1. Add two fields to the Patient table: InsuranceCompany and PolicyNumber. For patients with two policies, enter an additional record.
   2. Create a new table that includes all the data from the Patient table and the new fields, InsuranceCompany and PolicyNumber.
   3. Create a new Insurance table that includes the PatientNum, InsuranceCompany, and PolicyNumber fields. The Insurance table could have an autoincrementing key field or a compound key, PatientNum and PolicyNumber.
   4. Add two fields to the Session table: InsuranceCompany and PolicyNumber. For patients with two policies, enter an additional record.
2. A new table needs to be added to the Sports Physical Therapy database to track which sports the patients are playing. This data will be used to analyze the results of the therapies. Assuming that many patients play multiple sports, how would you design the table?
   1. SportsPlayed (PatientNum, SportName)
   2. SportsPlayed (PatientNum, SportName)
   3. SportsPlayed (SportsPlayedID, PatientNum, SportName)
   4. SportsPlayed (PatientNum, SportName1, SportName2, SportName3)
3. The Sports Physical Therapy clinic uses supplies when conducting therapies on patients. The database administrator is creating a table that lists all the possible supplies that the clinic orders. The design of this table is Supplies (SupplyID, Description, Price, SupplierID, SupplierName, SupplierAddress, SupplierPhone). Is this table designed correctly? If not, what normal form would correct it?
   1. no; first normal form
   2. no; second normal form
   3. no; third normal form
   4. yes, it is correctly designed.

Critical Thinking Questions

1. When the Sports Physical Therapy database was created, the database designers identified PatientNum and PatientSocialSecurityNumber as candidate keys for the Patient table. Why was PatientSocialSecurityNumber not chosen to be the key field of that table?

Answer: If the database were hacked, Social Security numbers would be released to criminals. Using a PatientNum minimizes the privacy risk to patients with information stored in the database.

1. What is the notation for each table in the Sports Physical Therapy database?

Answer: Patient (PatientNum, LastName, FirstName, Address, City, State, ZipCode, Balance); Therapist (TherapistID, LastName, FirstName, Street, City, State, ZipCode); Therapies (TherapyCode, Description, UnitOfTime); Session (SessionNum, SessionDate, PatientNum, LengthOfSession, TherapistID, TherapyCode)