Jared Beach

[Jmbeach1@crimson.ua.edu](mailto:Jmbeach1@crimson.ua.edu)

Dr. Susan Vrbsky

CS 457 – 001

Due Feb 12, 2016

Assignment 2

1.  What is the difference (if any) between the following - give definitions of each as well:

        a) program-data independence

* Program-data independence describes file processing in which a change to the structure of a file used by a program, does not require changing the programs that access the file as well.
  + Program-data independence is different from program-operation independence in that program-data independence is concerned with how data is stored and how it affects programs and program-operation independence is concerned with how operations are stored and re-used for the access and manipulation of the data.
  + Program-data independence is different from data independence in that program-data independence is concerned with how programs which use the data will be affected by changes to data-file structure and data independence is concerned with how upper levels of the database schema will be affected by lower-level changes to the schema.

        b) program-operation independence

* Program-operation independence describes a program in which operations are called by their interface (method signature) and the implementation of the operation is specified elsewhere. Thus, the program does not have to be changed if the operation implementation changes.
  + Program-operation independence is different from data independence in the sense that program-operation independence is concerned with how a program is affected by the changes to an operation for manipulating data and data independence is concerned with how the database schema is changed by changes to lower levels of the schema.

        c) data independence

* Data independence is the ability to change a database schema at a particular level of a database system without having to change the next higher level.

2.  The structure of a data model is described by what?

* The internal schema

3.   Using the [Company DB](http://cs457.cs.ua.edu/CompanyDBer.pdf) ER diagram, give an example of the following:

        a) an entity with more than one candidate key (list the keys)

* The department entity has candidate keys "Number" and "Name."

        b) a composite attribute

* The "Name" attribute of the "Employee" entity has subparts "Fname", "Minit", and "Lname."

        c) superkey that is not minimal

* The group of attributes {"SSN","Fname","Lname"} is a superkey in the "Employee" entity which is not minimal because if we remove "Fname" and "Lname" we still have a superkey.

        d) minimal superkey with more than one column

* The group of attributes {"Name", "Sex", "Birthdate", "Relationship"} is a superkey of the "Dependent" entity because none of the attributes are unique by themselves.

4. [Question 3.16](http://cs457.cs.ua.edu/2016S/Q3.16.pdf) in the textbook.

1. The combination of attributes "SecNo", "Sem", "Year", "CRoom", and "DaysTime" have to be unique in order to enforce that only one section can use a classroom at a particular "DaysTime" value during a specific semester and year.
2. The combination of attributes "Sem", "Year", "Teaches" (the attribute which specifies which teacher teaches the section), "SecNo", and "DaysTime" must be unique in order to enforce that an instructor can teach only one section at a particular DaysTime value during a specific semester and year.
3. The combination of attributes "Sem", "Year", "SecNo", and "Secs" (the attribute which specifies which course the section belongs to) must be unique in order to ensure that no section for the same course has the same course number as another section of the same course offered the same semester of the same year.

5. Which conditions must hold for a table to be a relation?  For each condition, explain why you think it was included in the relational model.

* The ordering of tuples (rows) in a table/relation does not matter. When retrieved, the tuples are usually displayed in a certain order, but it is not a requirement of a relation.
  + The condition that tuples must not be in a fixed order was included in order to give more flexibility in how they can be retrieved and in order to represent the tuples more factually by not assigning an arbitrary order to each one.
* A tuple or row contains values in a specific order.
  + This condition was probably included in order to simplify the storage of the data which must be in a particular order for retrieval from a disk.
* Each value in a tuple is atomic.
  + This enforces that if another relation is needed, it is not circumvented by making an attribute multi-valued. This makes the structure of a databse more logical and more extendable.
* The NULL value is used when a value is not known or does not apply to a tuple.
  + This ensures that false or fabricated data does not get inserted into a tuple for the sake of filling the attribute.
* The relation's schema makes assertions about its data. For example, a student relation could ensure that each student tuple must have a social security number.
  + This allows for more factual and realistic control over the entity being represented. Students, for example, in real life have a social security number, so making that assertion on the relation schema makes the data more accurate.

6. What are the advantages to using a minimal superkey in the relational model instead of one that is not minimal? Are there any disadvanatages?

* The advantages of using a minimal superkey compared to one that is not minimal are that it is less complicated because there are less attributes to have to group uniquely and that it is more easily implemented (unique id for example could be generated for any type of entity).
* The disadvantage of using a minimal superkey is that a minimal superkey might not match the entity's logical structure as well. For example, a CD entity might need an identifier along with a movie to be more uniquely identifiable, even if it could be unique with just the identifier.

7. Give a small example (different from any used in class) of 2 relation instances and illustrate the difference between the tuples resulting from a Cartesian product, an inner join and an outer join.

* Using the Company Employee and Project relations where Employee contains attributes SSN, Bdate, Name, Address, Salary, and Sex and Project contains attributes Name, Number, and Location. For example purposes, let's say that the Employee relation contains tuples:
  + (1111223333, 1/1/1973, John R. Smith, 1111 Lakeview Ln., $10, Male)
  + (2222334444, 1/1/1974, Sally S. Smith, 1131 Mountainview Ln., $11, Female)
* And the Project relation contains tuples:
  + (Build store, 1, 1234 Shopping Rd.)
  + (Build bank, 2, 1231 Banking Ln):
* A Cartesian product between the Company and Project relations would result in every single tuple from the Employee table being combined with every single tuple from the Project table:
  + (1111223333, 1/1/1973, John R. Smith, 1111 Lakeview Ln., $10, Male, Build store, 1, 1234 Shopping Rd.)
  + (1111223333, 1/1/1973, John R. Smith, 1111 Lakeview Ln., $10, Male, Build bank, 2, 1231 Banking Ln)
  + (2222334444, 1/1/1974, Sally S. Smith, 1131 Mountainview Ln., $11, Female, Build store, 1, 1234 Shopping Rd.)
  + (2222334444, 1/1/1974, Sally S. Smith, 1131 Mountainview Ln., $11, Female, Build bank, 2, 1231 Banking Ln)
* An inner join between the Company and product relations would result in only tuples where the tuples from the Employee relation have something in common with the tuples from the Project relation. For this example, assume that John R. Smith in the tuple from the Employee relation is working on project number 1 from the Project relation and that Sally S. Smith is working on project number 2:
  + (1111223333, 1/1/1973, John R. Smith, 1111 Lakeview Ln., $10, Male, Build store, 1, 1234 Shopping Rd.)
  + (2222334444, 1/1/1974, Sally S. Smith, 1131 Mountainview Ln., $11, Female, Build bank, 2, 1231 Banking Ln)
* An outer join will exclude data from the other relation which the first relation does not have anything in common. For this example, let's say that Sally S. Smith is not working on any project and that John R. Smith is still working on project 1. The result of selecting from the Employee relation left outer join the Project relation is as follows:
  + (1111223333, 1/1/1973, John R. Smith, 1111 Lakeview Ln., $10, Male, Build store, 1, 1234 Shopping Rd.)
  + (2222334444, 1/1/1974, Sally S. Smith, 1131 Mountainview Ln., $11, Female, NULL, NULL, NULL)