Database Normalization

Functional Dependencies

- A *Functional Dependency* describes a relationship between *attributes* within a single relation.
- An attribute is *functionally dependent* on another if we can use the value of one attribute to determine the value of another.
- Example: Employee_Name is functionally dependent on Social_Security_Number because Social_Security_Number can be used to uniquely determine the value of Employee_Name.
- We use the arrow symbol → to indicate a functional dependency.

  X → Y is read *X functionally determines Y*

  Here are a few more examples:

  Student_ID → Student_Major
  Student_ID, CourseNumber, Semester → Grade
  Course_Number, Section → Professor, Classroom, NumberOfStudents
  SKU → Compact_Disk_Title, Artist
  CarModel, Options, TaxRate → Car_Price

Keys and Uniqueness

- **Key**: One or more attributes that uniquely identify a tuple (row) in a relation
- The selection of keys will depend on the particular application being considered.
- In most cases the key for a relation will already be specified during the conversion from the E-R model to a set of relations.
- Users can also offer some guidance as to what would make an appropriate key.
- Recall that no two relations should have exactly the same values, thus a candidate key would consist of all of the attributes in a relation.
- A key functionally determines a tuple (row). So one functional dependency that can always be written is:

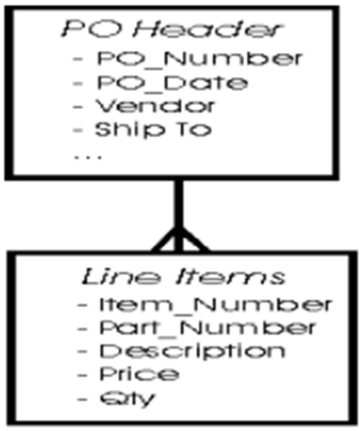
  The Key → All other attributes

Modification Anomalies

- Once our E-R model has been converted into relations, we may find that some relations are not properly specified. There can be a number of problems:

- **Deletion Anomaly**: Deleting one fact or data point from a relation results in other information being lost.
- **Insertion Anomaly**: Inserting a new fact or tuple into a relation requires we have information from two or more entities – this situation might not be feasible.
- **Update Anomaly**: Updating one fact in a relation requires us to update multiple tuples.
- Here is a quick example to illustrate these anomalies: A company has a Purchase Order form:

- Our dutiful consultant creates the E-R Model directly matching the purchase order:



When we follow the steps to convert to a set of relations this results in two relations (keys are underlined):

PO_HEADER (PO_Number, PODate, Vendor, Ship_To, ...)
LINE_ITEMS (PO_Number, ItemNum, PartNum, Description, Price, Qty)

Consider some sample data for the LINE_ITEMS relation:

PO_Number	ItemNum	PartNum	Description	Price	Qty
O101	I01	P99	Plate	$3.00	7
O101	I02	P98	Cup	$1.00	11
O101	I03	P77	Bowl	$2.00	6
O102	I01	P99	Plate	$3.00	5
O102	I02	P77	Bowl	$2.00	5
O103	I01	P33	Fork	$2.50	8

• What are some of the problems with this relation ?

1. What happens if we want to add the fact that Order O103 has quantity 5 of part P99 ?

2. What happens when we delete item I02 from Order O101 ?
3. What happens if we want to change the price of the Plate (P99)?
• These problems occur because the relation in question contains data about 2 or more *themes*.
• Typical way to solve these anomalies is to split the relation in to two or more relations – This is part of the *Process* called *Normalization* discussed next.

Normalization Process

Relations can fall into one or more categories (or classes) called *Normal Forms*
• **Normal Form**: A class of relations free from a certain set of modification anomalies.
• Normal forms are given names such as:

First normal form (1NF)
• Second normal form (2NF)
• Third normal form (3NF)
• Boyce-Codd normal form (BCNF)
• Fourth normal form (4NF)
• Fifth normal form (5NF)
• Domain-Key normal form (DK/NF)

These forms are cumulative. A relation in Third normal form is also in 2NF and 1NF.
• The *Normalization Process* for a given relation consists of:
• Specify the *Key* of the relation
• Specify the *functional dependencies* of the relation.
Sample data (tuples) for the relation can assist with this step.
• Apply the definition of each normal form (starting with 1NF).
• If a relation fails to meet the definition of a normal form, change the relation (most often by splitting the relation into two new relations) until it meets the definition.
• Re-test the modified/new relations to ensure they meet the definitions of each normal form.
In the next set of notes, each of the normal forms will be defined along with an example of the normalization step:

Un –Normalized:

Cust Name	Item	Shipping Address	Newsletter	Supplier	Supplier Phone	Price
Alan Smith	Xbox One	35 Palm st,Miami	Xbox News	Microsoft	(800) BUY-XBOX	250
Roger Banks	PlayStation 4	47 Campus Rd, Boston	PlayStation News	Sony	(800) BUY-SONY	300
Evan Wilson	Xbox One, PS Vita	28 Rock Av, Denver	Xbox News, PlayStation News	Wholesale	Toll Free	450
Alan Smith	PlayStation 4	47 Campus Rd, Boston	PlayStation News	Sony	(800) BUY-SONY	300

1st Normal Form:

- Each cell to be Single valued.

- Entried in a column are same type.
- Rows univequely identified-Add Unique ID, or Add more colums to make Unique.
- Note: The order of the rows and the order of the columns are irrelevant.

Cust ID	Cust Name	Item	Shipping Address	Newsletter	Supplier	Supplier Phone	Price
al_smith	Alan Smith	Xbox One	35 Palm st,Miami	Xbox News	Microsoft	(800) BUY-XBOX	250
Roger25	Roger Banks	PlayStation 4	47 Campus Rd, Boston	PlayStation News	Sony	(800) BUY-SONY	300
Wilson44	Evan Wilson	Xbox One,	28 Rock Av, Denver	Xbox News	Microsoft	(800) BUY-XBOX	250
Wilson44	Evan Wilson	PS Vita	28 Rock Av, Denver	PlayStation News	Sony	(800) BUY-SONY	200
	Alan Smith	PlayStation 4	47 Campus Rd, Boston	PlayStation News	Sony	(800) BUY-SONY	300

2nd Normal Form:

- All Attributes (Non-Key-Columns) dependent on the key.

Does price depend on Cust ID?

Cust ID	Cust Name	Item	Shipping Address	Newsletter	Supplier	Supplier Phone	Price
al_smith	Alan Smith	Xbox One	35 Palm st,Miami	Xbox News	Microsoft	(800) BUY-XBOX	250
Roger25	Roger Banks	PlayStation 4	47 Campus Rd, Boston	PlayStation News	Sony	(800) BUY-SONY	300
Wilson44	Evan Wilson	Xbox One,	28 Rock Av, Denver	Xbox News	Microsoft	(800) BUY-XBOX	250
Wilson44	Evan Wilson	PS Vita	28 Rock Av, Denver	PlayStation News	Sony	(800) BUY-SONY	200
Am_smith	Alan Smith	PlayStation 4	47 Campus Rd, Boston	PlayStation News	Sony	(800) BUY-SONY	300

2nd Normal form:

Cust ID	Cust Name	Shipping Address	Newsletter
al_smith	Alan Smith	35 Palm st,Miami	Xbox News
Roger25	Roger Banks	47 Campus Rd, Boston	PlayStation News
Wilson44	Evan Wilson	28 Rock Av, Denver	Xbox News
Wilson44	Evan Wilson	28 Rock Av, Denver	PlayStation News
Am_smith	Alan Smith	47 Campus Rd, Boston	PlayStation News

Item	Supplier	Supplier Phone	Price
Xbox One	Microsoft	(800) BUY-XBOX	250
PlayStation 4	Sony	(800) BUY-SONY	300
Xbox One,	Microsoft	(800) BUY-XBOX	250
PS Vita	Sony	(800) BUY-SONY	200
PlayStation 4	Sony	(800) BUY-SONY	300

Cust ID	Item
al_smith	Xbox One
Roger25	PlayStation 4
Wilson44	Xbox One
Wilson44	PS Vita
Am_smith	PlayStation 4

3rd Normal Form:

- All fields (column) can be determined Only by the Key in the table and no other column.

Cust ID	Cust Name	Shipping Address	Newsletter
al_smith	Alan Smith	35 Palm st,Miami	Xbox News
Roger25	Roger Banks	47 Campus Rd, Boston	PlayStation News
Wilson44	Evan Wilson	28 Rock Av, Denver	Xbox News
Wilson44	Evan Wilson	28 Rock Av, Denver	PlayStation News
Am_smith	Alan Smith	47 Campus Rd, Boston	PlayStation News

Item	Supplier	Price
Xbox One	Microsoft	250
PlayStation 4	Sony	300
PS Vita	Sony	200

Cust ID	Item
al_smith	Xbox One
Roger25	PlayStation 4
Wilson44	Xbox One
Wilson44	PS Vita
Am_smith	PlayStation 4

Supplier	Supplier Phone
Microsoft	(800) BUY-XBOX
Sony	(800) BUY-SONY

4th Normal Form:

- No Multivalued dependencies.

Cust ID	Cust Name	Shipping Address
al_smith	Alan Smith	35 Palm st,Miami
Roger25	Roger Banks	47 Campus Rd, Boston
Wilson44	Evan Wilson	28 Rock Av, Denver
Am_smith	Alan Smith	47 Campus Rd, Boston

Item	Supplier	Price
Xbox One	Microsoft	250
PlayStation 4	Sony	300
PS Vita	Sony	200

Cust ID	Item
al_smith	Xbox One
Roger25	PlayStation 4
Wilson44	Xbox One
Wilson44	PS Vita
Am_smith	PlayStation 4

Supplier	Supplier Phone
Microsoft	(800) BUY-XBOX
Sony	(800) BUY-SONY

Cust ID	Newsletter
al_smith	Xbox News
Roger25	PlayStation News
Wilson44	Xbox News
Wilson44	PlayStation News
Am_smith	PlayStation News