

# Module 4 Tutorial: Normalization

# Normalization

- ▶ Normalization is based on the observation that relations with certain properties are more effective in inserting, updating and deleting data than other sets of relations containing the same data
- ▶ Normalization is a multi-step process of decomposing unsatisfactory "bad" relations by breaking up their attributes into smaller relations

Source of the following hospital example: Atre, S. *Data Base: Structured Techniques for Design, Performance, and Management*.

# Unnormalized Relations

- ▶ First step in normalization is to convert the data into a two-dimensional table
- ▶ In unnormalized relations data can repeat within a column

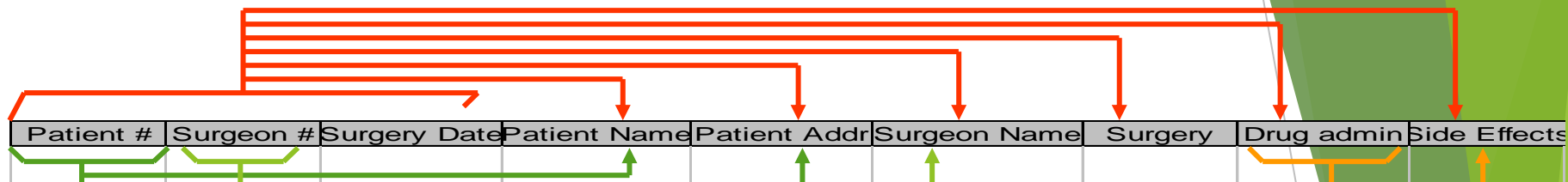
# Unnormalized Relation

Patient #	Surgeon #	Surg. date	Patient Name	Patient Addr	Surgeon	Surgery	Postop drug	ug side effect
1111	145 311	Jan 1, 1995; June 12, 1995	John White	15 New St. New York, NY	Beth Little Michael Diamond	Gallstone s removal; Kidney stones removal	Penicillin, none-	rash none
1234	243 467	Apr 5, 1994 May 10, 1995	Mary Jones	10 Main St. Rye, NY	Charles Field Patricia Gold	Eye Cataract removal Thrombos is removal	Tetracyclin e none	Fever none
2345	189	Jan 8, 1996	Charles Brown	Dogwood Lane Harrison, NY	David Rosen	Open Heart Surgery	Cephalosp orin	none
4876	145	Nov 5, 1995	Hal Kane	55 Boston Post Road, Chester, CN	Beth Little	Cholecyst ectomy	Demicillin	none
5123	145	May 10, 1995	Paul Kosher	Blind Brook Mamaronec k, NY	Beth Little	Gallstone s Removal	none	none
6845	243	Apr 5, 1994 Dec 15, 1984	Ann Hood	Hilton Road Larchmont, NY	Charles Field	Eye Cornea Replacem ent Eye cataract removal	Tetracyclin e	Fever

# First Normal Form

- ▶ To move to First Normal Form a relation must contain only atomic values at each row and column.
  - ▶ No repeating groups
  - ▶ A column or set of columns is called a Candidate Key when its values can uniquely identify the row in the relation.

# First Normal Form



Patient #	Surgeon #	Surgery Date	Patient Name	Patient Addr	Surgeon Name	Surgery	Drug admin	Side Effects
1111	145	01-Jan-95	John White	15 New St. New York, NY	Beth Little	Gallstone s removal	Penicillin	rash
1111	311	12-Jun-95	John White	15 New St. New York, NY	Michael Diamond	Kidney stones removal	none	none
1234	243	05-Apr-94	Mary Jones	10 Main St. Rye, NY	Charles Field	Eye Cataract removal	Tetracyclin e	Fever
1234	467	10-May-95	Mary Jones	10 Main St. Rye, NY	Patricia Gold	Thrombos is removal	none	none
2345	189	08-Jan-96	Charles Brown	Dogwood Lane Harrison, NY	David Rosen	Open Heart Surgery	Cephalosp orin	none
4876	145	05-Nov-95	Hal Kane	55 Boston Post Road, Chester, CN	Beth Little	Cholecyst ectomy	Demicillin	none
5123	145	10-May-95	Paul Kosher	Blind Brook Mamaronec k, NY	Beth Little	Gallstone s Removal	none	none
6845	243	05-Apr-94	Ann Hood	Hilton Road Larchmont, NY	Charles Field	Eye Cornea Replacem ent	Tetracyclin e	Fever
6845	243	15-Dec-84	Ann Hood	Hilton Road Larchmont, NY	Charles Field	Eye cataract removal	none	none

# 1NF Storage Anomalies


- ▶ **Insertion:** A new patient has not yet undergone surgery -- hence no surgeon # -- Since surgeon # is part of the key, we cannot insert.
- ▶ **Insertion:** If a surgeon is newly hired and has not operated yet -- there will be no way to include that person in the database.
- ▶ **Update:** If a patient comes in for a new procedure, and has moved, we need to change multiple address entries.
- ▶ **Deletion (type 1):** Deleting a patient record may also delete all info about a surgeon.
- ▶ **Deletion (type 2):** When there are functional dependencies (like side effects and drug) changing one item eliminates other information.

# Second Normal Form

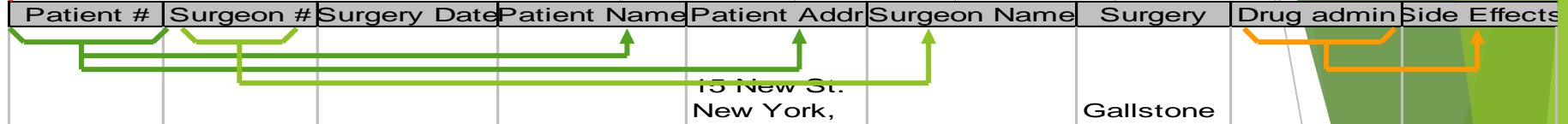
- ▶ A relation is said to be in Second Normal Form when every non-key attribute is **fully functionally dependent** on the primary key.
  - ▶ That is, every non-key attribute needs the full primary key for unique identification



# Why is this not in 2NF?



Patient #	Surgeon #	Surgery Date	Patient Name	Patient Addr	Surgeon Name	Surgery	Drug admin	Side Effects
1111	145	01-Jan-95	John White	15 New St. New York, NY	Beth Little	Gallstone s removal	Penicillin	rash
1111	311	12-Jun-95	John White	15 New St. New York, NY	Michael Diamond	Kidney stones removal	none	none
1234	243	05-Apr-94	Mary Jones	10 Main St. Rye, NY	Charles Field	Eye Cataract removal	Tetracyclin e	Fever
1234	467	10-May-95	Mary Jones	10 Main St. Rye, NY	Patricia Gold	Thrombos is removal	none	none
2345	189	08-Jan-96	Charles Brown	Dogwood Lane Harrison, NY	David Rosen	Open Heart Surgery	Cephalosp orin	none
4876	145	05-Nov-95	Hal Kane	55 Boston Post Road, Chester, CN	Beth Little	Cholecyst ectomy	Demicillin	none
5123	145	10-May-95	Paul Kosher	Blind Brook Mamaronec k, NY	Beth Little	Gallstone s Removal	none	none
6845	243	05-Apr-94	Ann Hood	Hilton Road Larchmont, NY	Charles Field	Eye Cornea Replacem ent	Tetracyclin e	Fever
6845	243	15-Dec-84	Ann Hood	Hilton Road Larchmont, NY	Charles Field	Eye cataract removal	none	none



# Second Normal Form

Patient #	Patient Name	Patient Address
1111	John White	15 New St. New York, NY
1234	Mary Jones	10 Main St. Rye, NY
2345	Charles Brown	Dogwood Lane Harrison, NY
4876	Hal Kane	55 Boston Post Road, Chester, Blind Brook
5123	Paul Kosher	Mamaroneck, NY
6845	Ann Hood	Hilton Road Larchmont, NY



# Second Normal Form

Surgeon #	Surgeon Name
145	Beth Little
189	David Rosen
243	Charles Field
311	Michael Diamond
467	Patricia Gold



# Second Normal Form

Patient #	Surgeon #	Surgery Date	Surgery	Drug Admin	Side Effects
1111	145	01-Jan-95	Gallstones removal	Penicillin	rash
1111	311	12-Jun-95	stones removal	none	none
1234	243	05-Apr-94	Eye Cataract removal	Tetracycline	Fever
1234	467	10-May-95	Thrombosis removal	none	none
2345	189	08-Jan-96	Open Heart Surgery	Cephalosporin	none
4876	145	05-Nov-95	Cholecystectomy	Demicillin	none
5123	145	10-May-95	Gallstones Removal	none	none
6845	243	15-Dec-84	Eye cataract removal	none	none
6845	243	05-Apr-94	Eye Cornea Replacement	Tetracycline	Fever

# 1NF Storage Anomalies Removed

- ▶ **Insertion:** Can now enter new patients without surgery.
- ▶ **Insertion:** Can now enter Surgeons who have not operated.
- ▶ **Deletion (type 1):** If Charles Brown dies, the corresponding tuples from Patient and Surgery tables can be deleted without losing information on David Rosen.
- ▶ **Update:** If John White comes in for third time, and has moved, we only need to change the Patient table

# 2NF Storage Anomalies

- ▶ **Insertion:** Cannot enter the fact that a particular drug has a particular side effect unless it is given to a patient.
- ▶ **Deletion:** If John White receives some other drug because of the penicillin rash, and a new drug and side effect are entered, we lose the information that penicillin can cause a rash
- ▶ **Update:** If drug side effects change (a new formula) we have to update multiple occurrences of side effects.

# Third Normal Form

- ▶ A relation is said to be in Third Normal Form if there is no transitive functional dependency between non-key attributes
  - ▶ When one non-key attribute can be determined with one or more non-key attributes there is said to be a transitive functional dependency.
- ▶ The side effect column in the Surgery table is determined by the drug administered
  - ▶ Side effect is transitively functionally dependent on drug so Surgery is not 3NF

# Why is this not in 3NF?


Diagram illustrating dependencies:

- Patient # and Surgeon # determine Surgery.
- Surgery determines Drug Admin and Side Effects.

Patient #	Surgeon #	Surgery Date	Surgery	Drug Admin	Side Effects
1111	145	01-Jan-95	Gallstones removal	Penicillin	rash
1111	311	12-Jun-95	stones removal	none	none
1234	243	05-Apr-94	Eye Cataract removal	Tetracycline	Fever
1234	467	10-May-95	Thrombosis removal	none	none
2345	189	08-Jan-96	Open Heart Surgery	Cephalosporin	none
4876	145	05-Nov-95	Cholecystectomy	Demicillin	none
5123	145	10-May-95	Gallstones Removal	none	none
6845	243	15-Dec-84	Eye cataract removal	none	none
6845	243	05-Apr-94	Eye Cornea Replacement	Tetracycline	Fever

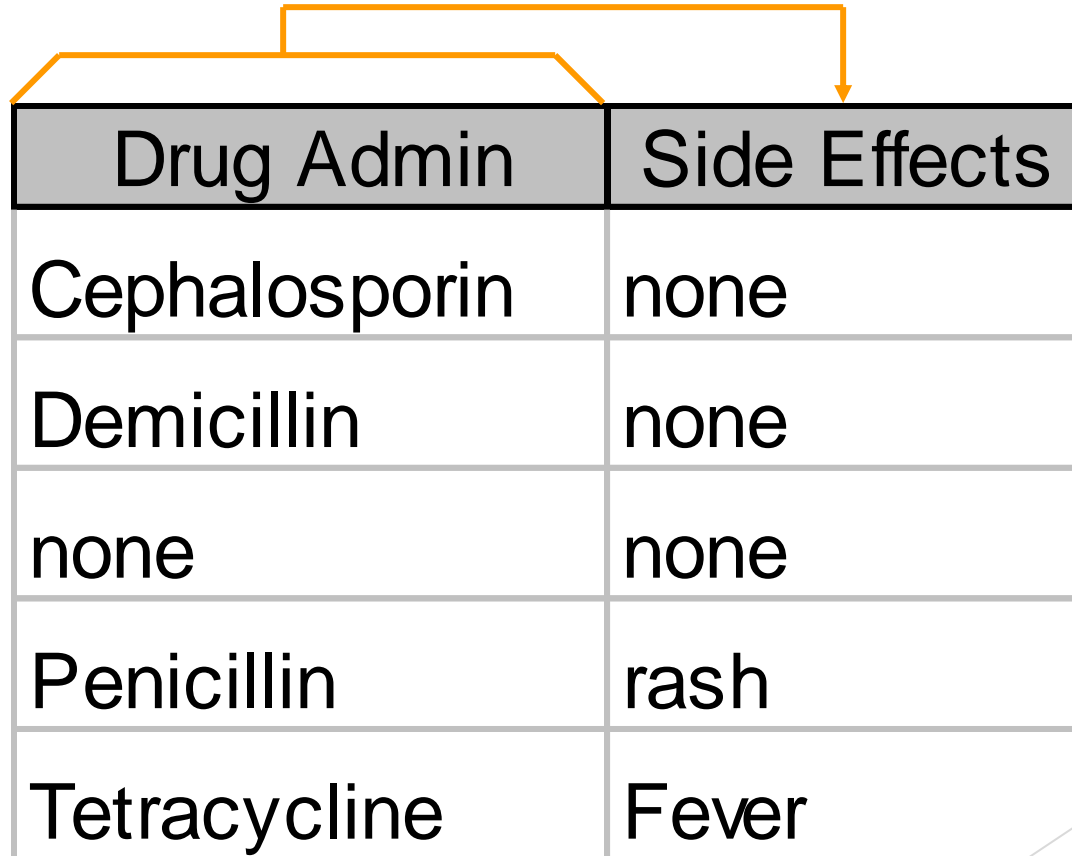


# Third Normal Form



Patient #	Surgeon #	Surgery Date	Surgery	Drug Admin
1111	145	01-Jan-95	Gallstones removal	Penicillin
1111	311	12-Jun-95	Kidney stones removal	none
1234	243	05-Apr-94	Eye Cataract removal	Tetracycline
1234	467	10-May-95	Thrombosis removal	none
2345	189	08-Jan-96	Open Heart Surgery	Cephalosporin
4876	145	05-Nov-95	Cholecystectomy	Demicillin
5123	145	10-May-95	Gallstones Removal	none
6845	243	15-Dec-84	Eye cataract removal	none
6845	243	05-Apr-94	Eye Cornea Replacement	Tetracycline

# Third Normal Form



Drug Admin	Side Effects
Cephalosporin	none
Demicillin	none
none	none
Penicillin	rash
Tetracycline	Fever

# 2NF Storage Anomalies Removed

- ▶ **Insertion:** We can now enter the fact that a particular drug has a particular side effect in the Drug relation.
- ▶ **Deletion:** If John White receives some other drug as a result of the rash from penicillin, the information on penicillin and rash is maintained.
- ▶ **Update:** The side effects for each drug appear only once.