

Component 4: Introduction to Information and Computer Science

Unit 6: Databases and SQL Lecture 4

This material was developed by Oregon Health & Science University, funded by the Department of Health and Human
Services, Office of the National Coordinator for Health Information Technology under Award Number 1U24OC000015.

Topic IV: Design a Simple Relational Database using Data Modeling and Normalization

- Description and Information Gathering
- Data Model
- Normalization, Functional Dependencies and Constraints
- Final design, Relationships, Primary keys and Foreign keys

Component 4/Unit 6-4

Health IT Workforce Curriculum
Version 2.0/Spring 2011

2

Description of Database

- Keep track of new medications that are in trial testing.
- Keep track of the medications, the trials for those medications and the clinical institutions doing the testing.

Component 4/Unit 6-4

Health IT Workforce Curriculum
Version 2.0/Spring 2011

3

Information Gathering

- Through meetings with users and looking at forms and reports it was determined that certain data about a clinical institution needed to be kept in the data base.
 - Name of the institution
 - Address
 - Contact information

Component 4/Unit 6-4

Health IT Workforce Curriculum
Version 2.0/Spring 2011

4

Information Gathering Continued

Keep track of medications and trials:

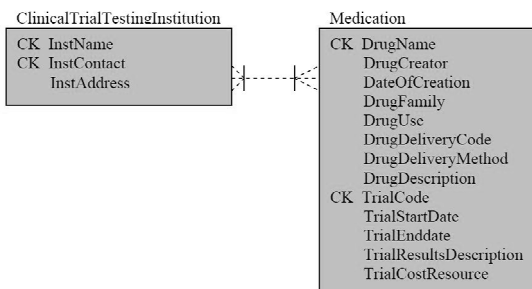
Drug name	Trial code
Drug creator	Trial start date
Date of Creation	Trial end date
Drug family	Trial results description
Drug use	Trial cost resource
Drug description	

Component 4/Unit 6-4

Health IT Workforce Curriculum
Version 2.0/Spring 2011

5

Data Model - First Attempt



Component 4/Unit 6-4

Health IT Workforce Curriculum
Version 2.0/Spring 2011

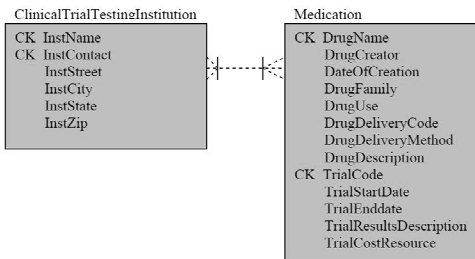
6

Normalization

- A database is normalized to eliminate data anomalies: Insert, Delete, Update
- Functional dependencies
- Constraints
 - Data rules that must be followed
- Referential Integrity Constraint

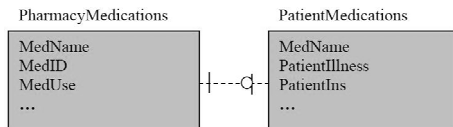
Functional Dependencies

- Data within a row can be shown to be dependent on a candidate key



Referential Integrity Constraint

- An attribute of one entity is a subset of an attribute of another entity.

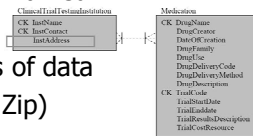


First Normal Form (1NF)

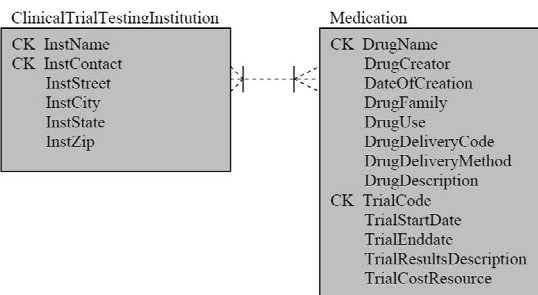
- Definition of a relation
 - Data rows within an entity must be unique and connected describe an instance of the entity (no data in a relation that is associated with something else).
 - Columns (attributes) are uniquely named, are of the same data type and will contain only one value.
 - The order of rows and columns is not important.

Putting the Example database in First Normal Form

- Nothing indicates that rows in the entities are not unique
- Attributes are connected to each entity
- Columns are uniquely named
- ClinicalTrialInstitution address contains pieces of data (Street, City, State and Zip)



Putting a Database In 1NF



Second Normal Form (2NF)

- 2NF eliminates deletion and insertion anomalies that are due to having an attribute or attributes dependent on something other than the key.
- This is especially true for composite keys.
- To be in second normal form attributes must be dependent on the whole key.

Component 4/Unit 6-4

Health IT Workforce Curriculum
Version 2.0/Spring 2011

13

Second Normal Form Continued

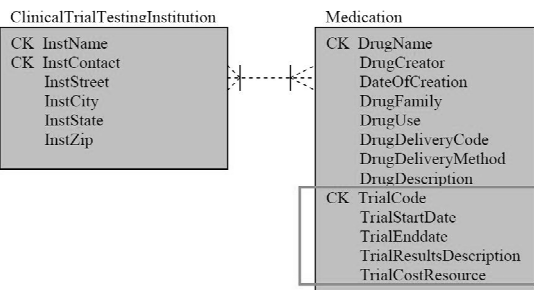
- A relation is in 2NF if all its non-key attributes are dependent on the entire key.
- A relation in 2NF must also be in 1NF.

Component 4/Unit 6-4

Health IT Workforce Curriculum
Version 2.0/Spring 2011

14

Putting The Trial DB In 2NF

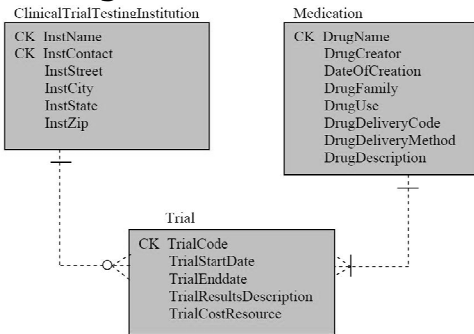


Component 4/Unit 6-4

Health IT Workforce Curriculum
Version 2.0/Spring 2011

15

Putting The Trial DB In 2NF



Component 4/Unit 6-4

Health IT Workforce Curriculum
Version 2.0/Spring 2011

16

Third Normal Form (3NF)

- 3NF eliminates deletion and insertion anomalies that are due to having an indirect dependency where an attribute is indirectly dependent on the key
- The attribute is directly dependent on an attribute that is dependent on the key
- The indirect dependency on the key is called a **transitive dependency**

Component 4/Unit 6-4

Health IT Workforce Curriculum
Version 2.0/Spring 2011

17

Third Normal Form Continued

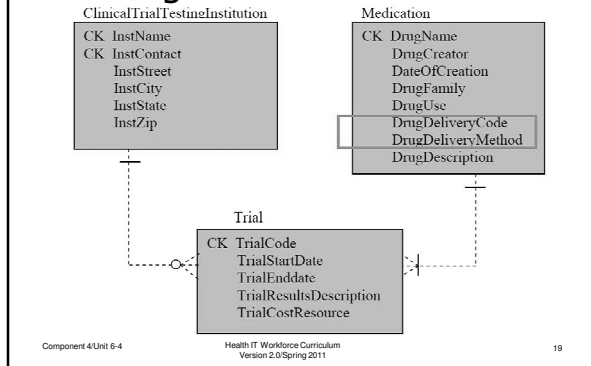
- A database is said to be 3NF if there are no transitive dependencies
- A database in 3NF must also be in 2NF and 1NF
- Many Database Administrators (DBAs) consider 3NF to be sufficient for most business and health care databases.
- Putting the database in a higher level of normalization may make the database less efficient.

Component 4/Unit 6-4

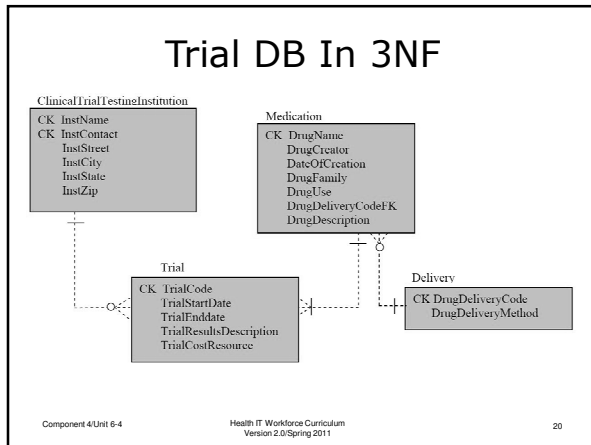
Health IT Workforce Curriculum
Version 2.0/Spring 2011

18

Putting The Trial DB In 3NF



Trial DB In 3NF



Other Normal Forms

- DBAs troubleshoot problems and on occasion will use normal forms beyond 3NF.
- A database can be de-normalized to solve some slow response problems.
- Boyce-Codd Normal Form (BCNF)
 - A determinant is an attribute that determines another attribute
 - A database is in Boyce-Codd form if every determinant is a candidate key

Other Normal Forms

- Fourth Normal Form (4NF)
 - This situation is rare
 - A multi-value dependency exists when there are a minimum of three attributes, two of the attributes are multi-valued and the values of the two multi-value attributes depend only on a 3rd attribute.
 - 4NF fixes an update anomaly that involves a multi-value dependency
 - A database is in 4NF when there are no multi-value dependencies

Component 4/Unit 6-4

Health IT Workforce Curriculum
Version 2.0/Spring 2011

22

Other Normal Forms

- Fifth Normal Form (5NF) or Project-Join Normal Form (PJNF)
 - Extremely rare
 - Generalization of multi-valued dependencies
 - Difficult to deal with
- Domain Key Normal Form (DKNF)
 - Generalization of other non-time constraints
 - Difficult to deal with

Component 4/Unit 6-4

Health IT Workforce Curriculum
Version 2.0/Spring 2011

23

Evolution of the Data Model

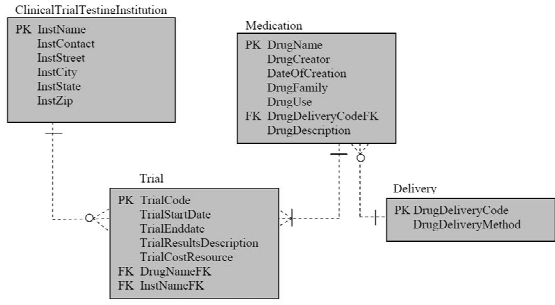
- Data model progresses from being volatile with many changes to a database design with little change or surprises
- In the final design, entities become tables, relationships show minimum and maximum cardinality and primary/foreign keys are chosen

Component 4/Unit 6-4

Health IT Workforce Curriculum
Version 2.0/Spring 2011

24

Final Design of the Trial DB



Component 4/Unit 6-4

Health IT Workforce Curriculum
Version 2.0/Spring 2011

25
