## Normalisation to 3NF

Database Systems Lecture 11 Natasha Alechina

## In This Lecture

- Normalisation to 3NF
  - Data redundancy
  - Functional dependencies
  - Normal forms
  - · First, Second, and Third Normal Forms
- For more information
  - Connolly and Begg chapter 13
  - Ullman and Widom ch.3.6.6 (2<sup>nd</sup> edition),
  - 3.5 (3<sup>rd</sup> edition)

### **Redundancy and Normalisation**

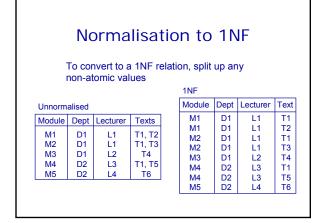
- Redundant data
  - Can be determined from other data in the database
  - Leads to various
  - problems
  - INSERT anomalies
     UPDATE anomalies
  - DELETE anomalies
  - DELETE anomalies

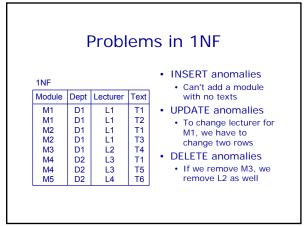
## Normalisation Aims to reduce data

- redundancy
- Redundancy is expressed in terms of dependencies
- Normal forms are defined that do not
- have certain types of dependency

## First Normal Form

- In most definitions of the relational model
  - All data values should be atomic
- This means that table entries should be single values, not sets or composite objects
- A relation is said to be in first normal form (1NF) if all data values are atomic





## **Functional Dependencies**

- Redundancy is often caused by a functional dependency
- A functional dependency (FD) is a link between two sets of attributes in a relation
- We can normalise a relation by removing undesirable FDs

• A set of attributes, A, functionally determines another set, B, or: there exists a functional dependency between A and B ( $A \rightarrow B$ ), if whenever two rows of the relation have the same values for all the attributes in A, then they also have the same values for all the attributes in B.

## Example

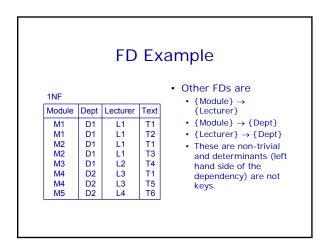
- {ID, modCode}  $\rightarrow$  {First, Last, modName}
- $\{modCode\} \rightarrow \{modName\}$
- $\{ID\} \rightarrow \{First, Last\}$

ID	First	Last	modCode	modName
111	Joe	Bloggs	G51PRG	Programming
222	Anne	Smith	G51DBS	Databases

#### FDs and Normalisation · We define a set of Not all FDs cause a 'normal forms' problem We identify various sorts of FD that do Each normal form has fewer FDs than the last Each normal form Since FDs represent removes a type of FD redundancy, each that is a problem normal form has less · We will also need a redundancy than the way to remove FDs last

#### **Properties of FDs** • In any relation Rules for FDs The primary key FDs any set of attributes in that relation • Reflexivity: If *B* is a subset of *A* then $A \rightarrow B$ $K \rightarrow X$ Augmentation: If K is the primary key, X is a set of $A \rightarrow B$ then $A \cup C \rightarrow B \cup C$ attributes Transitivity: · Same for candidate If $A \to B$ and $B \to C$ then $A \to C$ keys Any set of attributes is FD on itself $X \rightarrow X$

		FD	EX	ample
1NF				• The primary key is {Module, Text} so
Module	Dept	Lecturer	Text	{Module, Text} $\rightarrow$
M1	D1	L1	T1	{Dept, Lecturer}
M1	D1	L1	T2	<ul> <li>'Trivial' FDs, eq:</li> </ul>
M2	D1	L1	T1	${\text{Text, Dept}} \rightarrow {\text{Text}}$
M2	D1	L1	T3	
M3	D1	L2	T4	${Module} \rightarrow {Module}$
M4	D2	L3	T1	{Dept, Lecturer} $\rightarrow$ {
M4	D2	L3	T5	
M5	D2	L4	T6	



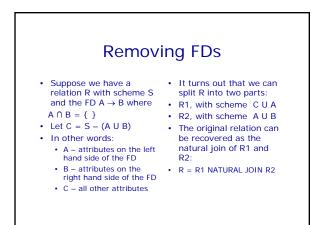
## Partial FDs and 2NF

- Partial FDs:
  - A FD, A → B is a partial FD, if some attribute of A can be removed and the FD still holds
     Formally, there is some
  - proper subset of A,  $C \subset A$ , such that  $C \rightarrow B$ Let us call attributes
- which are part of some candidate key, key attributes, and the rest non-key attributes.
- Second normal form: • A relation is in second
- normal form (2NF) if it is in 1NF and no non-key attribute is partially dependent on a candidate key
- In other words, no C → B where C is a strict subset of a candidate key and B is a non-key attribute.

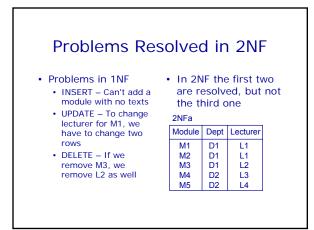
## Second Normal Form

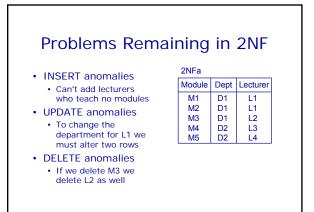
Module	Dept	Lecturer	Text
M1	D1	L1	T1
M1	D1	L1	T2
M2	D1	L1	T1
M2	D1	L1	T3
M3	D1	L2	T4
M4	D2	L3	T1
M4	D2	L3	T5
M5	D2	L4	T6

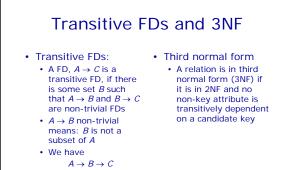
#### 



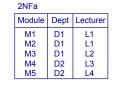
Module	Dept	Lecturer	Text	Module	Dept	Lecturer	Module	Тех
M1	D1	L1	T1	M1	D1	L1	M1	T1
M1	D1		T2	M2	D1	L1	M1	T2
M2	D1		T1	M3	D1		M2	T1
M2	D1	L1	T3	M4	D2	L3	M2	Т3
M3	D1	L2	T4	M5	D2	L4	M3	T4
M4	D2	L3	T1				M4	T1
M4	D2	L3	T5				M4	T5
M5	D2	L4	T6				M1	Т6





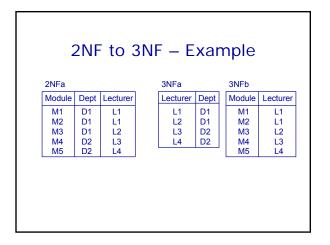


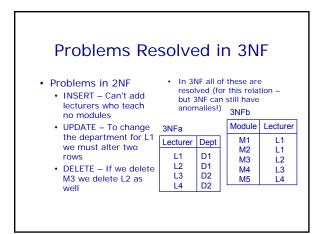
## Third Normal Form

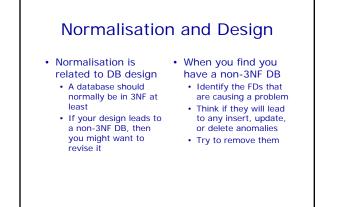


# 2NFa is not in 3NF We have the FDs {Module} → {Lecturer} {Lecturer} → {Dept} So there is a

transitive FD from the primary key {Module} to {Dept}







## Next Lecture

- More normalisation
  - Lossless decomposition; why our reduction to 2NF and 3NF is lossless
  - Boyce-Codd normal form (BCNF)
  - Higher normal forms
  - Denormalisation
- For more information
  - · Connolly and Begg chapter 14
  - Ullman and Widom chapter 3.6