

Normalisation

- Main objective in database design is to create an accurate and efficient representation of the data, its relationships, and constraints
- To do this we identify a suitable set of relations (tables)

Data Redundancy

- Aim - to group attributes into relations to minimise redundancy and reduce space required
- Problems associated with data redundancy are illustrated by comparing the emp, dept and grade relation with the employee relations

EMPLOYEE										
EMPNO	ENAME	JOB	MGR	DEPTNO	DNAME	LOC	SAL	GRADE	LOSAL	HISAL
405	MARCH	ADMIN	938	2	RESEARCH	YORK	18000	1	17000	21999
535	BYRNE	SALES	734	3	SALES	BIRMINGHAM	26000	3	24000	29999
557	BELL	SALES	734	3	SALES	BIRMINGHAM	22500	2	22000	23999
602	BIRD	MANAGER	875	2	RESEARCH	YORK	39750	4	30000	49999
690	AHMAD	SALES	734	3	SALES	BIRMINGHAM	22500	2	22000	23999
734	COX	MANAGER	875	3	SALES	BIRMINGHAM	38500	4	30000	49999
818	POLLARD	MANAGER	875	1	ACCOUNTING	LONDON	34500	4	30000	49999
824	REES	ANALYST	602	2	RESEARCH	YORK	40000	4	30000	49999
875	PARKER	PRESIDENT		1	ACCOUNTING	LONDON	60000	5	50000	99999
880	TURNER	SALES	734	3	SALES	BIRMINGHAM	25000	3	24000	29999
912	HAYES	ADMIN	824	2	RESEARCH	YORK	21000	1	17000	21999
936	CASSY	ADMIN	734	3	SALES	BIRMINGHAM	19500	1	17000	21999
938	GIBSON	ANALYST	602	2	RESEARCH	YORK	40000	4	30000	49999
970	BLACK	ADMIN	818	1	ACCOUNTING	LONDON	23000	2	22000	23999

EMP							
EMPNO	ENAME	JOB	MGR	HIREDATE	SAL	COMM	DEPTNO
405	MARCH	ADMIN	938	13/06/1997	18000		2
535	BYRNE	SALES	734	15/08/1997	26000	300	3
557	BELL	SALES	734	26/03/2000	22500	500	3
602	BIRD	MANAGER	875	31/10/1997	39750		2
690	AHMAD	SALES	734	05/12/1997	22500	1400	3
734	COX	MANAGER	875	11/06/2002	38500		3
818	POLLARD	MANAGER	875	14/05/2000	34500		1
824	REES	ANALYST	602	05/03/2000	40000		2
875	PARKER	PRESIDENT		09/07/2002	60000		1
880	TURNER	SALES	734	04/06/2001	25000	0	3
912	HAYES	ADMIN	824	04/06/2001	21000		2
936	CASSY	ADMIN	734	23/07/2002	19500		3
938	GIBSON	ANALYST	602	05/12/1997	40000		2
970	BLACK	ADMIN	818	21/11/1997	23000		1

DEPT			GRADE		
DEPTNO	DNAME	LOC	GRADE	LOSAL	HISAL
1	ACCOUNTING	LONDON	1	17000	21999
2	RESEARCH	YORK	2	22000	23999
3	SALES	BIRMINGHAM	3	24000	29999
4	OPERATIONS	LEEDS	4	30000	49999
			5	50000	99999

Data Redundancy

EMPLOYEE										
EMPNO	ENAME	JOB	MGR	DEPTNO	DNAME	LOC	SAL	GRADE	LOSAL	HISAL
405	MARCH	ADMIN	938	2	RESEARCH	YORK	18000	1	17000	21999
535	BYRNE	SALES	734	3	SALES	BIRMINGHAM	26000	3	24000	29999
557	BELL	SALES	734	3	SALES	BIRMINGHAM	22500	2	22000	23999
602	BIRD	MANAGER	875	2	RESEARCH	YORK	39750	4	30000	49999
690	AHMAD	SALES	734	3	SALES	BIRMINGHAM	22500	2	22000	23999
734	COX	MANAGER	875	3	SALES	BIRMINGHAM	38500	4	30000	49999
818	POLLARD	MANAGER	875	1	ACCOUNTING	LONDON	34500	4	30000	49999
824	REES	ANALYST	602	2	RESEARCH	YORK	40000	4	30000	49999
875	PARKER	PRESIDENT		1	ACCOUNTING	LONDON	60000	5	50000	99999
880	TURNER	SALES	734	3	SALES	BIRMINGHAM	25000	3	24000	29999
912	HAYES	ADMIN	824	2	RESEARCH	YORK	21000	1	17000	21999
936	CASSY	ADMIN	734	3	SALES	BIRMINGHAM	19500	1	17000	21999
938	GIBSON	ANALYST	602	2	RESEARCH	YORK	40000	4	30000	49999
970	BLACK	ADMIN	818	1	ACCOUNTING	LONDON	23000	2	22000	23999

- Clearly, many data items repeated
- A 'better' structure is possible with **three** tables

Data Redundancy

EMP							
EMPNO	ENAME	JOB	MGR	HIREDATE	SAL	COMM	DEPTNO
405	MARCH	ADMIN	938	13/06/1997	18000		2
535	BYRNE	SALES	734	15/08/1997	26000	300	3
557	BELL	SALES	734	26/03/2000	22500	500	3
602	BIRD	MANAGER	875	31/10/1997	39750		2
690	AHMAD	SALES	734	05/12/1997	22500	1400	3
734	COX	MANAGER	875	11/06/2002	38500		3
818	POLLARD	MANAGER	875	14/05/2000	34500		1
824	REES	ANALYST	602	05/03/2000	40000		2
875	PARKER	PRESIDENT		09/07/2002	60000		1
880	TURNER	SALES	734	04/06/2001	25000	0	3
912	HAYES	ADMIN	824	04/06/2001	21000		2
936	CASSY	ADMIN	734	23/07/2002	19500		3
938	GIBSON	ANALYST	602	05/12/1997	40000		2
970	BLACK	ADMIN	818	21/11/1997	23000		1

DEPT			GRADE		
DEPTNO	DNAME	LOC	GRADE	LOSAL	HISAL
1	ACCOUNTING	LONDON	1	17000	21999
2	RESEARCH	YORK	2	22000	23999
3	SALES	BIRMINGHAM	3	24000	29999
4	OPERATIONS	LEEDS	4	30000	49999
			5	50000	99999

- Reduces redundancy, decreases storage requirements
- Also other advantages - reduces the chance of update anomalies

Update Anomalies

- Relations that contain redundant information may potentially suffer from update anomalies.
- 3 types of update anomalies
 - Modification
 - Deletion
 - Insertion

Update Anomalies

Exercise - for the **employee** relation implement these changes

- Department 3 moves to Edinburgh
- Staff member 875 leaves
- Add the missing department

Update Anomalies

Now do the same for the **emp, dept** and **grade** relations implement these changes

- Department 3 moves to Edinburgh
- Staff member 875 leaves
- Add the missing department

Appropriate Design

- Clearly some designs are 'better' than others
- 'Better' here means
 - Appropriate
 - Efficient
- Normalisation is a process of generating a better design through 'decomposition'

EMPLOYEE										
EMPNO	ENAME	JOB	MGR	DEPTNO	DNAME	LOC	SAL	GRADE	LOSAL	HISAL
405	MARCH	ADMIN	938	2	RESEARCH	YORK	18000	1	17000	21999
535	BYRNE	SALES	734	3	SALES	BIRMINGHAM	26000	3	24000	29999
557	BELL	SALES	734	3	SALES	BIRMINGHAM	22500	2	22000	23999
602	BIRD	MANAGER	875	2	RESEARCH	YORK	39750	4	30000	49999
690	AHMAD	SALES	734	3	SALES	BIRMINGHAM	22500	2	22000	23999
734	COX	MANAGER	875	3	SALES	BIRMINGHAM	38500	4	30000	49999
818	POLLARD	MANAGER	875	1	ACCOUNTING	LONDON	34500	4	30000	49999
824	REES	ANALYST	602	2	RESEARCH	YORK	40000	4	30000	49999
875	PARKER	PRESIDENT		1	ACCOUNTING	LONDON	60000	5	50000	99999
880	TURNER	SALES	734	3	SALES	BIRMINGHAM	25000	3	24000	29999
912	HAYES	ADMIN	824	2	RESEARCH	YORK	21000	1	17000	21999
936	CASSY	ADMIN	734	3	SALES	BIRMINGHAM	19500	1	17000	21999
938	GIBSON	ANALYST	602	2	RESEARCH	YORK	40000	4	30000	49999
970	BLACK	ADMIN	818	1	ACCOUNTING	LONDON	23000	2	22000	23999

decomposition

EMP							
EMPNO	ENAME	JOB	MGR	HIREDATE	SAL	COMM	DEPTNO
405	MARCH	ADMIN	938	13/06/1997	18000		2
535	BYRNE	SALES	734	15/08/1997	26000	300	3
557	BELL	SALES	734	26/03/2000	22500	500	3
602	BIRD	MANAGER	875	31/10/1997	39750		2
690	AHMAD	SALES	734	05/12/1997	22500	1400	3
734	COX	MANAGER	875	11/06/2002	38500		3
818	POLLARD	MANAGER	875	14/05/2000	34500		1
824	REES	ANALYST	602	05/03/2000	40000		2
875	PARKER	PRESIDENT		09/07/2002	60000		1
880	TURNER	SALES	734	04/06/2001	25000		3
912	HAYES	ADMIN	824	04/06/2001	21000	0	2
936	CASSY	ADMIN	734	23/07/2002	19500		3
938	GIBSON	ANALYST	602	05/12/1997	40000		2
970	BLACK	ADMIN	818	21/11/1997	23000		1

DEPT			GRADE		
DEPTNO	DNAME	LOC	GRADE	LOSAL	HISAL
1	ACCOUNTING	LONDON	1	17000	21999
2	RESEARCH	YORK	2	22000	23999
3	SALES	BIRMINGHAM	3	24000	29999
4	OPERATIONS	LEEDS	4	30000	49999
			5	50000	99999

Normalisation

- A technique for producing a set of relations with desirable properties, given the data requirements of an enterprise
- Developed by E.F. Codd (1972)
- Often performed as a series of tests on a relation to determine whether it satisfies or violates the requirements of a given **normal form** (NF)

Lossless-join and Dependency Preservation Properties

There are properties that should hold in decompositions

Lossless-join property

Dependency preservation property

Lossless-join and Dependency Preservation Properties

Lossless-join property enables us to find any instance of the original relation from corresponding instances in the smaller relations

Lossless-Join decomposition (LJD)

- Can the original information be reformed (typically with a join)?

EMP	ENAME	JOB	MGR	HIREDATE	SAL	COMM	DEPTNO
405	MARCH	ADMIN	938	13/06/1997	18000		2
535	BYRNE	SALES	734	15/08/1997	26000	300	3
557	BELL	SALES	734	26/03/2000	22300	500	3
602	BIRD	MANAGER	875	11/10/1997	39750		2
690	AHMAD	SALES	734	05/12/1997	22500	1400	3
734	COX	MANAGER	875	11/06/2002	38500		3
818	POLLARD	MANAGER	875	14/05/2000	34500		1
824	REES	ANALYST	602	05/03/2000	40000		2
875	PARKER	PRESIDENT		09/07/2002	60000		1
880	TURNER	SALES	734	04/06/2001	25000	0	3
912	HAYES	ADMIN	824	04/06/2001	21000		2
936	CASSY	ADMIN	734	23/07/2002	19500		3
938	GIBSON	ANALYST	602	05/12/1997	40000		2
970	BLACK	ADMIN	818	21/11/1997	23000		1

DEPT	DNAME	LOC	GRADE	LOSAL	HISAL
1	ACCOUNTING	LONDON	1	17000	21999
2	RESEARCH	YORK	2	22000	23999
3	SALES	BIRMINGHAM	3	24000	29999
4	OPERATIONS	LEEDS	4	30000	49999
			5	50000	99999

EMPLOYEE											
EMPNO	ENAME	JOB	MGR	DEPTNO	DNAME	LOC	SAL	GRADE	LOSAL	HISAL	
405	MARCH	ADMIN	938	2	RESEARCH	YORK	18000	1	17000	21999	
535	BYRNE	SALES	734	3	SALES	BIRMINGHAM	26000	3	24000	29999	
557	BELL	SALES	734	3	SALES	BIRMINGHAM	22500	2	22000	23999	
602	BIRD	MANAGER	875	2	RESEARCH	YORK	39750	4	30000	49999	
690	AHMAD	SALES	734	3	SALES	BIRMINGHAM	22500	2	22000	23999	
734	COX	MANAGER	875	3	SALES	BIRMINGHAM	38500	4	30000	49999	
818	POLLARD	MANAGER	875	1	ACCOUNTING	LONDON	34500	4	30000	49999	
824	REES	ANALYST	602	2	RESEARCH	YORK	40000	4	30000	49999	
875	PARKER	PRESIDENT		1	ACCOUNTING	LONDON	60000	5	50000	99999	
880	TURNER	SALES	734	3	SALES	BIRMINGHAM	25000	3	24000	29999	
912	HAYES	ADMIN	824	2	RESEARCH	YORK	21000	1	17000	21999	
936	CASSY	ADMIN	734	3	SALES	BIRMINGHAM	19500	1	17000	21999	
938	GIBSON	ANALYST	602	2	RESEARCH	YORK	40000	4	30000	49999	
970	BLACK	ADMIN	818	1	ACCOUNTING	LONDON	23000	2	22000	23999	

Lossless-Join decomposition (LJD)

- Can the original information be reformed (typically with a join)?
- Consider these decompositions - are they LJDs?

empno, ename, job, mgr, dname, grade, loc, dname

grade, losal, hisal, deptno

empno, ename, job, mgr, deptno, grade

deptno, dname, loc

grade, losal, hisal

Lossless-Join decomposition (LJD)

- Consider these decompositions - are they LJDs?

empno, ename, job, mgr, deptno

deptno, dname, loc

grade, losal, hisal

empno, deptno, sal

empno, ename, job, mgr, dname, grade, loc

deptno, dname

grade, losal, hisal

Dependency Preserving Decomposition (DPD)

- There is a relationship between the two attributes **empno** and **ename**
- called a **functional dependency**

EMPNO	ENAME
405	MARCH
535	BYRNE
557	BELL
602	BIRD
690	AHMAD
734	COX
818	POLLARD
824	REES
875	PARKER
880	TURNER
912	HAYES
936	CASSY
938	GIBSON
970	BLACK

Dependency Preserving Decomposition (DPD)

- If **empno** and **ename** are placed in two separate tables alone, the relationship is lost (and is not recoverable)
- A **DPD** must retain the (required) functional dependencies

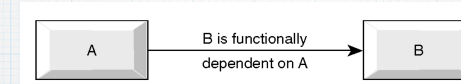
EMPNO
405
535
557
602
690
734
818
824
875
880
912
936
938
970

ENAME
MARCH
BYRNE
BELL
BIRD
AHMAD
COX
POLLARD
REES
PARKER
TURNER
HAYES
CASSY
GIBSON
BLACK

Functional Dependency

- Describes the relationship between attributes in a relation
- A and B are attributes of relation R, B is functionally dependent on A ($A \rightarrow B$), if each value of A in R is associated with exactly one value of B in R

Functional Dependency



A property of the meaning or semantics of the attributes in a relation

The determinant of a functional dependency refers to the attribute or group of attributes on the left-hand side of the arrow

Example - Functional Dependency

empno \rightarrow ename

We say "ename is functionally dependant on empno"

557 \rightarrow bell

An empno is associated with only one ename

EMPNO	ENAME
405	MARCH
535	BYRNE
557	BELL
602	BIRD
690	AHMAD
734	COX
818	POLLARD
824	REES
875	PARKER
880	TURNER
912	HAYES
936	CASSY
938	GIBSON
970	BLACK
340	BELL

Example - Functional Dependency

ename \nrightarrow empno

empno is NOT functionally dependant on ename

bell \rightarrow 557
bell \rightarrow 340

An ename is associated with more than one empno

EMPNO	ENAME
405	MARCH
535	BYRNE
557	BELL
602	BIRD
690	AHMAD
734	COX
818	POLLARD
824	REES
875	PARKER
880	TURNER
912	HAYES
936	CASSY
938	GIBSON
970	BLACK
340	BELL

Functional Dependency

Syntax

$\text{empno} \longrightarrow \text{ename, job}$

ename OR job is functionally dependant on empno

$\text{empno, deptno} \longrightarrow \text{job}$

job is functionally dependant on empno AND dept no

Functional Dependency

Syntax

Armstrongs Axioms show how FDs can be combined and split apart

eg $A \longrightarrow B, C$

Would be the same as

eg $A \longrightarrow B$

$A \longrightarrow C$

Not done here - 3rd year!

Exercise

EMPLOYEE										
EMPNO	ENAME	JOB	MGR	DEPTNO	DNAME	LOC	SAL	GRADE	LOSAL	HISAL
405	MARCH	ADMIN	938	2	RESEARCH	YORK	18000	1	17000	21999
535	BYRNE	SALES	734	3	SALES	BIRMINGHAM	26000	3	24000	29999
557	BELL	SALES	734	3	SALES	BIRMINGHAM	22500	2	22000	23999
602	BIRD	MANAGER	875	2	RESEARCH	YORK	39750	4	30000	49999
690	AHMAD	SALES	734	3	SALES	BIRMINGHAM	22500	2	22000	23999
734	COX	MANAGER	875	3	SALES	BIRMINGHAM	38500	4	30000	49999
818	POLLARD	MANAGER	875	1	ACCOUNTING	LONDON	34500	4	30000	49999
824	REES	ANALYST	602	2	RESEARCH	YORK	40000	4	30000	49999
875	PARKER	PRESIDENT		1	ACCOUNTING	LONDON	60000	5	50000	99999
880	TURNER	SALES	734	3	SALES	BIRMINGHAM	25000	3	24000	29999
912	HAYES	ADMIN	824	2	RESEARCH	YORK	21000	1	17000	21999
936	CASSY	ADMIN	734	3	SALES	BIRMINGHAM	19500	1	17000	21999
938	GIBSON	ANALYST	602	2	RESEARCH	YORK	40000	4	30000	49999
970	BLACK	ADMIN	818	1	ACCOUNTING	LONDON	23000	2	22000	23999

- Identify functional dependencies in the employee relation
- What is the relationship between the number of attributes and number of FDs?

The Process of Normalisation

- Formal technique for analyzing a relation based on its primary key and the functional dependencies between the attributes of that relation
- Often executed as a series of steps. Each step corresponds to a specific normal form, which has known properties

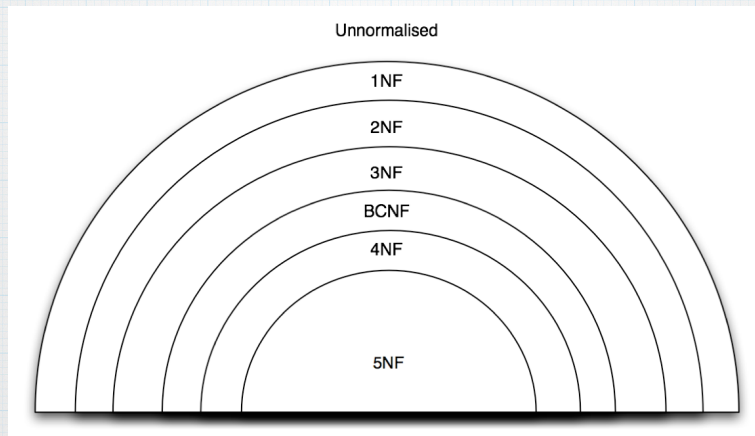
Normalisation

- Four most commonly used normal forms are first (1NF), second (2NF), third (3NF) and Boyce-Codd (BCNF) normal forms
- Based on functional dependencies among the attributes of a relation
- A relation can be normalised to a specific form to prevent the possible occurrence of update anomalies

The Process of Normalisation

- As normalisation proceeds, we say that the relations become progressively more restricted (stronger)
- They also become less vulnerable to update anomalies

Relationship Between Normal Forms



Unnormalised Form (UNF)

- A table that contains one or more repeating groups
- To create an unnormalised table
 - Transform the data from the information source (e.g. form) into table format with columns and rows
- Often trivial

Unnormalised example

<div> <div>Page 1</div> <div> DreamHome Customer Rental Details </div> <div>Date 7-Oct-98</div> </div>						
<div> <div>Customer Name John Kay</div> <div>Customer Number CR76</div> </div>						
Property Number	Property Address	Rent Start	Rent Finish	Rent	Owner Number	Owner Name
PG4	6 Lawrence St, Glasgow	1-Jul-94	31-Aug-96	350	CO40	Tina Murphy
PG16	5 Novar Dr, Glasgow	1-Sep-96	1-Sep-98	450	CO93	Tony Shaw

Unnormalised table

Customer No	Cname	Property No	Paddress	RentStart	RentFinish	Rent	Owner No	Oname
CR76	John Kay	PG4	6 Lawrence St, Glasgow	1-Jul-94	31-Aug-96	350	CO40	Tina Murphy
		PG16	5 Novar Dr, Glasgow	1-Sep-96	1-Sep-98	450	CO93	Tony Shaw
CR56	Aline Stewart	PG4	6 Lawrence St, Glasgow	1-Sep-92	10-Jun-94	350	CO40	Tina Murphy
		PG36	2 Manor Rd, Glasgow	10-Oct-94	1-Dec-95	375	CO93	Tony Shaw
		PG16	5 Novar Dr, Glasgow	1-Jan-96	10-Aug-96	450	CO93	Tony Shaw

First Normal Form (1NF)

- A relation in which the intersection of each row and column contains one and only one value

UNF to 1NF

- Nominate an attribute or group of attributes to act as the key for the unnormalised table
- Identify the repeating group(s) in the unnormalised table which repeats for the key attribute(s)

UNF to 1NF

- Remove the repeating group by
 - Entering appropriate data into the empty columns of rows containing the repeating data ('flattening' the table)
 - Or by
 - Placing the repeating data along with a copy of the original key attribute(s) into a separate relation

UNF to 1NF

Customer No	Cname	Property No	Address	RentStart	RentFinish	Rent	Owner No	Oname
CR76	John Kay	PG4	6 Lawrence St, Glasgow	1-Jul-94	31-Aug-96	350	CO40	Tina Murphy
		PG16	5 Novar Dr, Glasgow	1-Sep-96	1-Sep-08	450	CO93	Tony Shaw
CR56	Aline Stewart	PG4	6 Lawrence St, Glasgow	1-Sep-92	10-Jun-94	350	CO40	Tina Murphy
		PG36	2 Manor Rd, Glasgow	10-Oct-94	1-Dec-95	375	CO93	Tony Shaw
		PG16	5 Novar Dr, Glasgow	1-Jan-96	10-Aug-96	450	CO93	Tony Shaw

Customer No	Property No	Cname	Address	RentStart	RentFinish	Rent	Owner No	Oname
CR76	PG4	John Kay	6 Lawrence St, Glasgow	1-Jul-94	31-Aug-96	350	CO40	Tina Murphy
CR76	PG16	John Kay	5 Novar Dr, Glasgow	1-Sep-96	1-Sep-08	450	CO93	Tony Shaw
CR56	PG4	Aline Stewart	6 Lawrence St, Glasgow	1-Sep-92	10-Jun-94	350	CO40	Tina Murphy
CR56	PG36	Aline Stewart	2 Manor Rd, Glasgow	10-Oct-94	1-Dec-95	375	CO93	Tony Shaw
CR56	PG16	Aline Stewart	5 Novar Dr, Glasgow	1-Jan-96	10-Aug-96	450	CO93	Tony Shaw

UNF to 1NF

- Sometimes it can be difficult
- Have a look at DEPT and GRADE

Exercise

Customer No	Property No	Cname	Address	RentStart	RentFinish	Rent	Owner No	Oname
CR76	PG4	John Kay	6 Lawrence St, Glasgow	1-Jul-94	31-Aug-96	350	CO40	Tina Murphy
CR76	PG16	John Kay	5 Novar Dr, Glasgow	1-Sep-96	1-Sep-08	450	CO93	Tony Shaw
CR56	PG4	Aline Stewart	6 Lawrence St, Glasgow	1-Sep-92	10-Jun-94	350	CO40	Tina Murphy
CR56	PG36	Aline Stewart	2 Manor Rd, Glasgow	10-Oct-94	1-Dec-95	375	CO93	Tony Shaw
CR56	PG16	Aline Stewart	5 Novar Dr, Glasgow	1-Jan-96	10-Aug-96	450	CO93	Tony Shaw

- Identify Primary key for this table
- Redefine the term Primary Key in terms of functional dependencies

Second Normal Form (2NF)

- Based on the concept of full functional dependency
- Full functional dependency indicates that if
 - A and B are attributes of a relation,
 - B is fully dependent on A if B is functionally dependent on A but not on any proper subset of A

Second Normal Form (2NF)

- A relation that is in 1NF and every non-primary-key attribute is fully functionally dependent on the primary key

You may have $AB \rightarrow C$

but do you in fact have $B \rightarrow C$?

1NF to 2NF

- Identify the functional dependencies in the relation
- If partial dependencies exist on the primary key remove them by placing them in a new relation along with a copy of their determinant

1NF to 2NF

Customer No	Property No	Cname	Paddress	RentStart	RentFinish	Rent	Owner_No	Oname
CR76	PG4	John Kay	6 Lawrence St, Glasgow	1-Jul-94	31-Aug-96	350	CO40	Tina Murphy
CR76	PG16	John Kay	5 Novar Dr, Glasgow	1-Sep-96	1-Sep-98	450	CO93	Tony Shaw
CR56	PG4	Aline Stewart	6 Lawrence St, Glasgow	1-Sep-92	10-Jun-94	350	CO40	Tina Murphy
CR56	PG36	Aline Stewart	2 Manor Rd, Glasgow	10-Oct-94	1-Dec-95	375	CO93	Tony Shaw
CR56	PG16	Aline Stewart	5 Novar Dr, Glasgow	1-Jan-96	10-Aug-96	450	CO93	Tony Shaw

- CNo, PNo is primary key
- But not all attributes are dependant on both
- Which are dependant on CNo?
- Which are dependant on PNo alone?

1NF to 2NF

Rental			
Customer No	Property No	RentStart	RentFinish
CR76	PG4	1-Jul-94	31-Aug-96
CR76	PG16	1-Sep-96	1-Sep-08
CR56	PG4	1-Sep-92	10-Jun-94
CR56	PG36	10-Oct-94	1-Dec-95
CR56	PG16	1-Jan-96	10-Aug-96

Customer	
Customer No	Cname
CR76	John Kay
CR56	Aline Stewart

Property				
Property No	Address	Rent	Owner_No	Oname
PG4	6 Lawrence St, Glasgow	350	CO40	Tina Murphy
PG36	2 Manor Rd, Glasgow	375	CO93	Tony Shaw
PG16	5 Novar Dr, Glasgow	450	CO93	Tony Shaw

- Check for DPD and LJD at this stage

Third Normal Form (3NF)

- Based on the concept of transitive dependency
- Transitive Dependency is a condition where A, B and C are attributes of a relation such that
 - if $A \rightarrow B$ and $B \rightarrow C$ then C is transitively dependent on A through B
- (Provided that A is not functionally dependent on B or C)

Third Normal Form (3NF)

- A relation that is in 1NF and 2NF and in which no non-primary-key attribute is transitively dependent on the primary key

You may have $A \rightarrow C$

but do you in fact have $A \rightarrow B$ and $B \rightarrow C$?

2NF to 3NF

- Identify the primary key in the 2NF relation
- Identify functional dependencies in the relation
- If transitive dependencies exist on the primary key remove them by placing them in a new relation along with a copy of their determinant

2NF to 3NF

Rental			
Customer No	Property No	RentStart	RentFinish
CR76	PG4	1-Jul-94	31-Aug-96
CR76	PG16	1-Sep-96	1-Sep-08
CR56	PG4	1-Sep-92	10-Jun-94
CR56	PG36	10-Oct-94	1-Dec-95
CR56	PG16	1-Jan-96	10-Aug-96

Customer	
Customer No	Cname
CR76	John Kay
CR56	Aline Stewart

Property				
Property No	Address	Rent	Owner No	Owner
PG4	6 Lawrence St, Glasgow	350	CO40	Tina Murphy
PG36	2 Manor Rd, Glasgow	375	CO93	Tony Shaw
PG16	5 Novar Dr, Glasgow	450	CO93	Tony Shaw

- Check for transitive dependencies in Customer, Property and Rental

2NF to 3NF

- Property is split

Property				
Property No	Address	Rent	Owner No	Owner
PG4	6 Lawrence St, Glasgow	350	CO40	Tina Murphy
PG36	2 Manor Rd, Glasgow	375	CO93	Tony Shaw
PG16	5 Novar Dr, Glasgow	450	CO93	Tony Shaw

Property			
Property No	Address	Rent	Owner No
PG4	6 Lawrence St, Glasgow	350	CO40
PG36	2 Manor Rd, Glasgow	375	CO93
PG16	5 Novar Dr, Glasgow	450	CO93

Owner	
Owner No	Owner
CO40	Tina Murphy
CO93	Tony Shaw

- Again - check for LJD and DPD

Normalisation Summary of 3NF Relations

Customer	
Customer No	Cname
CR76	John Kay
CR56	Aline Stewart

Rental			
Customer No	Property No	RentStart	RentFinish
CR76	PG4	1-Jul-94	31-Aug-96
CR76	PG16	1-Sep-96	1-Sep-08
CR56	PG4	1-Sep-92	10-Jun-94
CR56	PG36	10-Oct-94	1-Dec-95
CR56	PG16	1-Jan-96	10-Aug-96

Property			
Property No	Address	Rent	Owner No
PG4	6 Lawrence St, Glasgow	350	CO40
PG36	2 Manor Rd, Glasgow	375	CO93
PG16	5 Novar Dr, Glasgow	450	CO93

Owner	
Owner No	Owner
CO40	Tina Murphy
CO93	Tony Shaw

Exercise

- Normalise COMPLAINTS

Exercise

- Normalise PROJ MAN