Normalization

Normalization

- Main objective in developing a logical data model for relational database systems is to create an accurate representation of the data, its relationships, and constraints.
- To achieve this objective, we must identify a suitable set of relations => The Normalization Technique
- Developed by E.F. Codd (1972).

Normalization

- Often performed as a series of tests on a relation to determine whether it satisfies or violates the requirements of a given normal form.
- 4 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 normalized to a specific form to prevent the possible occurrence of update anomalies.

Data Redundancy

- Major aim of relational database design is to group attributes into relations to minimize data redundancy and reduce file storage space required by base relations.

Staff Relation

Staff_No	*SName*	*SAddress*	*Position*	*Salary*	*Branch_No*
SL21	John White	19 Taylor St, London	Manager	30000	B5
SG37	Ann Beech	81 George St, Glasgow	Snr Asst	12000	B3
SG14	David Ford	63 Ashby St, G			
SA9	Mary Howe	2 Elm Pl, Aber			
SG5	Susan Brand	5 Gt Western R			
SL41	Julie Lee	28 Malvern St,			

Branch Relation

Branch_No	*BAddress*	*Tel_No*
B5	22 Deer Rd, London	0171-886-1212
B7	16 Argyll St, Aberdeen	01224-67125
B3	163 Main St, Glasgow	0141-339-2178

Data Redundancy

Staff_Branch Relation

Staff_No	*SName*	*SAddress*	*Position*	*Salary*	*Branch_No*	*BAddress*	*Tel_No*
SL21	John White	19 Taylor St, London	Manager	30000	B5	22 Deer Rd, London	0171-886-1212
SG37	Ann Beech	81 George St, Glasgow	Snr Asst	12000	B3	163 Main St, Glasgow	0141-339-2178
SG14	David Ford	63 Ashby St, Glasgow	Deputy	18000	B3	163 Main St, Glasgow	0141-339-2178
SA9	Mary Howe	2 Elm Pl, Aberdeen	Assistant	9000	B7	16 Argyll St, Aberdeen	01224-67125
SG5	Susan Brand	5 Gt Western Rd, Glasgow	Manager	24000	B3	163 Main St, Glasgow	0141-339-2178
SL41	Julie Lee	28 Malvern St, Kilburn	Assistant	9000	B5	22 Deer Rd, London	0171-886-1212

- **Staff_Branch relation has redundant data; the details of a branch are repeated for every member of staff.**

Update Anomalies

- **Relations that contain redundant information may potentially suffer from update anomalies.**
- **Types of update anomalies include**
  - **Insertion**
  - **Deletion (eg. in the Staff_Branch relation, for a particular branch, removal of the last member of staff => lost of branch information)**
  - **Modification**

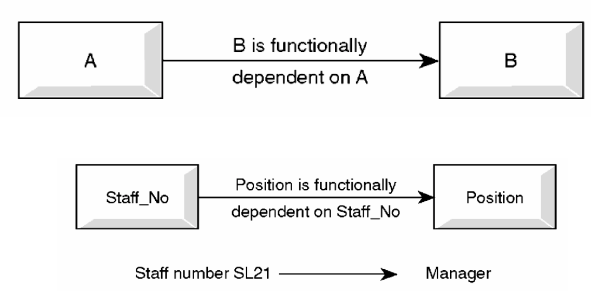
Functional Dependency

Functional Dependency:

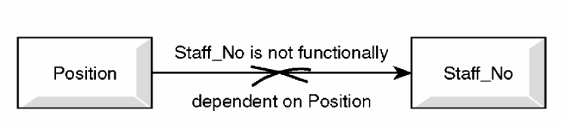
- **Main concept associated with normalization.**
- **Functional Dependency**
  - **Describes the relationship between attributes in a relation.**
  - **For example, if A and B are attributes of relation R, B is functionally dependent on A (denoted A -> B), if each value of A in R is associated with exactly one value of B in R.**

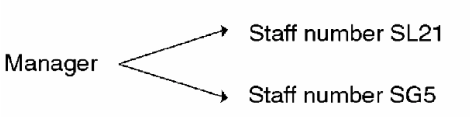
Functional Dependency

Diagrammatic representation:



Functional Dependency





3 Types of Functional Dependencies

Full Dependency

- **In a relation, the attribute(s) B is fully functional dependent on A if B is functionally dependent on A, but not on any proper subset of A.**

Partial Dependency

- **If there is some attribute that can be removed from A and the dependency still holds.**

Eg. Staff_No, Sname -> Branch_No

3 Types of Functional Dependencies

Transitive Dependency

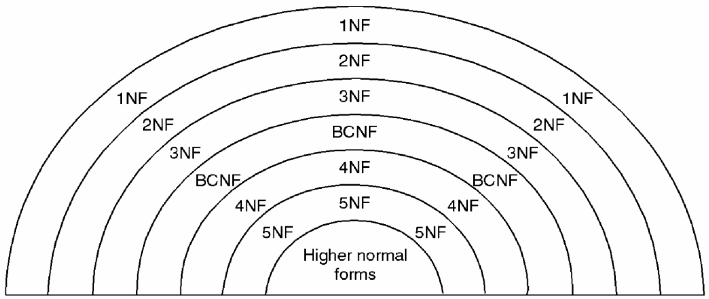
- **In a relation, if attribute(s) A->B and B->C, then C is transitively dependent on A via B (provided that A is not functionally dependent on B or C)**

Eg. Staff_No->Branch_No and Branch_No->BAddress

The Process of Normalization

- **Normalization: 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.**
- **As normalization proceeds, the relations become progressively more restricted (stronger) in format and also less vulnerable to update anomalies.**

Relationship Between Normal Forms



Unnormalized Normal Form

- **Unnormalized Normal Form (UNF)**
  - **A table that contains one or more repeating groups.**
  - **To create an unnormalized table**

    Transform the data from the information source (e.g. form) into table format with columns and rows.

Example - A Report

Page 1 **DreamHome** Customer Rental Details Date 7-Oct-98

Customer Name *John Kay* Customer Number *CR76*

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

Example - UNF

Customer_Rental 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-June-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.

Converting from UNF to 1NF:

- **Select attribute(s) to act as the key.**
- **Identify the repeating group(s) in the unnormalized table which repeats for the key attribute(s).**
- **Remove the repeating group by**
  - **Entering data into empty columns of rows which contain the repeating data.**

  Or by
  - **Placing the repeating data along with a copy of the original key attribute(s) into a separate relation.**

Example - Normalization UNF to 1NF

Customer_Rental 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-June-94	350	CO40	Tina Murphy
		PG36	2 Manor Rd, Glasgow	10-Oct-94	1-Dec-95	375	CO93	Tony Shaw
		PG16	5 Novar Dr,	1-Jan-96	10-Aug-96	450	CO93	Tony Shaw

Customer_Rental Relation

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

Example - UNF to 1NF (Alternative)

Customer_Rental 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-June-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

Prop_Rental_Owner Relation

Customer_No	*Property_No*	*PAddress*	*RentStart*	*RentFinish*	*Rent*	*Owner_No*	*OName*
CR76	PG4	6 Lawrence St, Glasgow	1-Jul-94	31-Aug-96	350	CO40	Tina Murphy
CR76	PG16	5 Novar Dr, Glasgow	1-Sep-96	1-Sep-98	450	CO	
CR56	PG4	6 Lawrence St, Glasgow	1-Sep-92	10-Jun-94	350	CO	
CR56	PG36	2 Manor Rd, Glasgow	10-Oct-94	1-Dec-95	375	CO	
CR56	PG16	5 Novar Dr, Glasgow	1-Jan-96	10-Aug-96	450	CO	

Customer Relation

Customer_No	*CName*
CR76	John Kay
CR56	Aline Stewart

Second Normal Form

Based on the concept of full functional dependency.

A 2NF relation is in 1NF and every non-primary-key attribute is fully functionally dependent on the primary key.

Converting from 1NF to 2NF:

- **Identify the primary key for the 1NF relation.**
- **Identify the functional dependencies in the relation.**
- **If partial dependencies exist on the primary key remove them by placing then in a new relation along with a copy of their determinant.**

FDs for Customer_Rental Relation

Customer_No	*Property_No*	*CName*	*PAddress*	*RentStart*	*RentFinish*	*Rent*	*Owner_No*	*OName*

Primary key: Customer_No + Property_No

Full Functional Dependency:
(Customer_No+Property_No)->(RentStart, RentFinish)

Partial Dependency:
(Customer_No+Property_No)->Cname
(Customer_No+Property_No)->(Paddress, Rent, Owner_No, Oname)

Customer_Rental to 2NF Relations

Customer_Rental Relation

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							
CR56	PG4							
CR56	PG36							
CR56	PG16							

Customer Relation

Customer_No	*CName*
CR76	John Kay
CR56	Aline Stewart

Rental Relation

Customer_No	*Property_No*	*RentStart*	*RentFinish*
CR76	PG4	1-Jul-94	31-Aug-96
CR76	PG16	1-Sep-96	1-Sep-98
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_Owner Relation

Property_No	*PAddress*	*Rent*	*Owner_No*	*OName*
PG4	6 Lawrence St, Glasgow	350	CO40	Tina Murphy
PG16	5 Novar Dr, Glasgow	450	CO93	Tony Shaw
PG36	2 Manor Rd, Glasgow	375	CO93	Tony Shaw

Third Normal Form (3NF)

Based on the concept of transitive dependency.

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

Converting from 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 dominant.**

Property_Owner to 3NF Relations

Property_Owner Relation

Property_No	*PAddress*	*Rent*	*Owner_No*	*OName*

Transitive Dependency:
(Customer_No+Property_No)->Owner_No
Owner_No ->OName

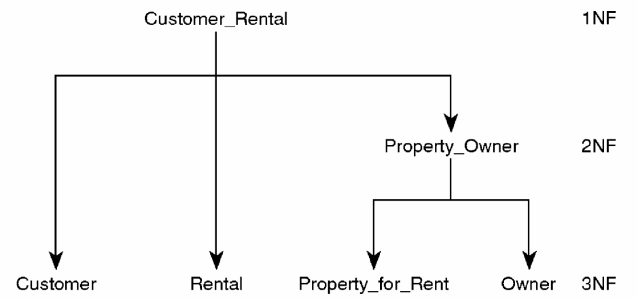
Property_for_Rent Relation

Property_No	*PAddress*	*Rent*	*Owner_No*
PG4	6 Lawrence St, Glasgow	350	CO40
PG16	5 Novar Dr, Glasgow	450	CO93
PG36	2 Manor Rd, Glasgow	375	CO93

Owner Relation

Owner_No	*OName*
CO40	Tina Murphy
CO93	Tony Shaw

Process of Decomposition



Summary of 3NF Relations

Customer Relation

Customer_No	*CName*
CR76	John Kay
CR56	Aline Stewart

Rental Relation

Customer_No	*Property_No*	*RentStart*	*RentFinish*
CR76	PG4	1-Jul-94	31-Aug-96
CR76	PG16	1-Sep-96	1-Sep-98
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_for_Rent Relation

Property_No	*PAddress*	*Rent*	*Owner_No*
PG4	6 Lawrence St, Glasgow	350	CO40
PG16	5 Novar Dr, Glasgow	450	CO93
PG36	2 Manor Rd, Glasgow	375	CO93

Owner Relation

Owner_No	*OName*
CO40	Tina Murphy
CO93	Tony Shaw

Boyce-Codd Normal Form (BCNF)

- **Based on functional dependencies that takes into account all candidate keys in a relation.**
- **For a relation with only one candidate key, 3NF and BCNF are equivalent.**
- **A relation is in BCNF, if and only if every determinant is a candidate key.**
- **Violation of BCNF may occur in a relation that**
  - **contains 2 (or more) composite keys**
  - **which overlap and share at least 1 attribute**

3NF to BCNF

- **Identify all candidate keys in the relation.**
- **Identify all functional dependencies in the relation.**
- **If functional dependencies exists in the relation where their determinants are not candidate keys for the relation, remove the functional dependencies by placing them in a new relation along with a copy of their determinant.**

Example - 3NF to BCNF Relations

Client_Interview Relation

Client_No	*Interview_Date*	*Interview_Time*	*Staff_No*	*Room_No*
CR76	13-May-98	10.30	SG5	G101
CR56	13-May-98	12.00	SG5	G101
CR74	13-May-98	12.00	SG37	G102
CR56	1-Jul-98	10.30	SG5	G102

Everyday a staff is allocated a room only. Staff may share rooms. Clients maybe interviewed again.

(Client_No, Interview_Date) -> (Interview_Time, Staff_No, Room_No)
(Staff_No, Interview_Date, Interview_Time) -> Client_No
(Room_No, Interview_date, Interview_Time) -> Staff_No, Client_No
(Staff_No, Interview_Date) -> Room_No

Example - 3NF to BCNF Relations

Client_Interview Relation

Client_No	*Interview_Date*	*Interview_Time*	*Staff_No*	*Room_No*
CR76	13-May-98	10.30	SG5	G101
CR56	13-May-98	12.00	SG5	G101
CR74	13-May-98	12.00	SG37	G102
CR56	1-Jul-98	10.30	SG5	G102

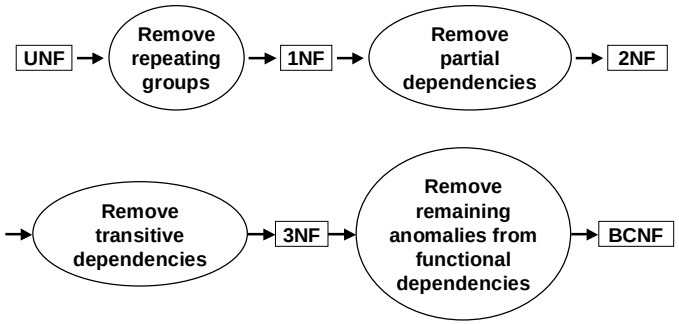
Interview Relation

Client_No	*Interview_Date*	*Interview_Time*	*Staff_No*
CR76	13-May-98	10.30	SG5
CR56	13-May-98	12.00	SG5
CR74	13-May-98	12.00	SG37
CR56	1-Jul-98	10.30	SG5

Staff_Room Relation

Staff_No	*Interview_Date*	*Room_No*
SG5	13-May-98	G101
SG37	13-May-98	G102
SG5	1-Jul-98	G102

The Process of Normalization up to BCNF



Example

Page 1 — ***DreamHome*** Property Inspection Report — Date 1-Oct-98

Property Number PG4 — Property Address 6 Lawrence St, Glasgow

Inspection Date	Inspection Time	Comments	Staff Number	Staff Name	Car Reg
18-Oct-96	10.00	Need to replace crockery	SG37	Ann Beech	M231 JGR
22-Apr-97	09.00	In good order	SG14	David Ford	M533 HDR
1-Oct-98	12.00	Damp rot in bathroom	SG14	David Ford	N721 HFR

Example - UNF to 1NF Relation

Property_Inspection Table

Property_No	*PAddress*	*IDate*	*ITime*	*Comments*	*Staff_No*	*SName*	*Car_Reg*
PG4	6 Lawrence St, Glasgow	18-Oct-96	10.00	need to replace crockery	SG37	Ann Beech	M231 JGR
		22-Apr-97	09.00	in good order	SG14	David Ford	M533 HDR
		1-Oct-98	12.00	damp rot in bathroom	SG14	David Ford	N721 HFR
PG16	5 Novar Dr, Glasgow	22-Apr-96	13.00	replace living room carpet	SG14	David Ford	M533 HDR
		24-Oct-97	14.00	good condition	SG37	Ann Beech	N721 HFR

Property_Inspection Relation

Property_No	*IDate*	*ITime*	*PAddress*	*Comments*	*Staff_No*	*SName*	*Car_Reg*
PG4	18-Oct-96	10.00	6 Lawrence St, Glasgow	need to replace crockery	SG37	Ann Beech	M231 JGR
PG4	22-Apr-97	09.00	6 Lawrence St, Glasgow	in good order	SG14	David Ford	M533 HDR
PG4	1-Oct-98	12.00	6 Lawrence St, Glasgow	damp rot in bathroom	SG14	David Ford	N721 HFR
PG16	22-Apr-96	13.00	5 Novar Dr, Glasgow	replace living room carpet	SG14	David Ford	M533 HDR
PG16	24-Oct-97	14.00	5 Novar Dr, Glasgow	good condition	SG37	Ann Beech	N721 HFR

Example - 1NF to 2NF

1NF: Property_Inspection (Property_No, IDate, ITime, Paddress, Comments, Staff_No, Sname, Car_Reg)

Full Functional Dependency:

(Property_No+IDate)->(ITime, Comments, Staff_No, Sname, Car_Reg)

Partial Dependency:
(Property_No+IDate)->(PAddress)

=> 2NF:
Prop (Property_No, Paddress)
Prop_Inspection (Property_No, IDate, ITime, Comments, Staff_No, Sname, Car_Reg)

Example - 2NF to 3NF

Transitive Dependency in Prop_Inspect:

(Property_No+IDate) -> Staff_No

Staff_No -> Sname

=> 3NF:
Staff (Staff_No, Sname)
Prop_Inspection (Property_No, IDate, ITime, Comments, Staff_No, Car_Reg)

Example - 3NF to BCNF

Prop (Property_No,Paddress)
Staff (Staff_No, Sname)
Prop_Inspection (Property_No, IDate, ITime, Comments, Staff_No, Car_Reg)

Prop and Staff are already in BCNF.

FDs of Prop_Inspect:

- **(Property_No, IDate)->(ITime, Comments, Staff_No, Car_Reg)**
- **(Staff_No, Idate) -> Car_Reg**
- **(Car_Reg, Idate, ITime) -> (Property_No, Comments, Staff_No)**
- **(Staff_No, Idate, ITime) -> (Property_No, Comments)**

Example – BCNF

Prop (Property_No,Paddress)

Staff (Staff_No, Sname)

Inspection (Property_No, IDate, ITime, Comments, Staff_No)

Staff_Car (Staff_No, IDate, Car_Reg)

Process of Decomposition