CSC 443 – Database Management Systems

Lecture 5 – Normalization

Normalization

- We want our database to be a clear representation of the data, its relationships and constraints
- We can identify relationship using a technique called ***normalization***.
- Normalization is a bottom-up technique where we examine the relationship between attributes and reconfigure the tables accordingly.

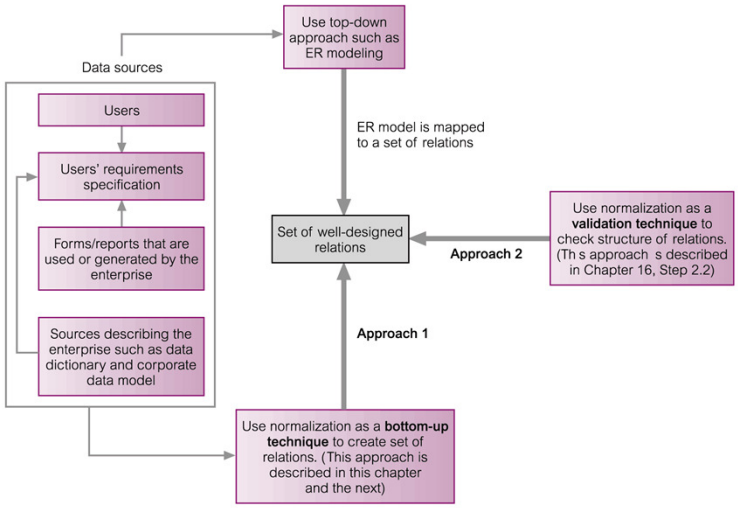
Purpose of Normalization

- Characteristics of a suitable set of relations include:
  - the minimal number of attributes necessary to support the data requirements of the enterprise;
  - attributes with a close logical relationship are found in the same relation;
  - minimal redundancy with each attribute represented only once with the important exception of attributes that form all or part of foreign keys.

Purpose of Normalization

- The benefits of using a database that has a suitable set of relations is that the database will be:
  - easier for the user to access and maintain the data;
  - take up minimal storage space on the computer.

How Normalization Supports Database Design



Data Redundancy and Update Anomalies

- Major aim of relational database design is to group attributes into relations to minimize data redundancy.

Data Redundancy and Update Anomalies

- Potential benefits for implemented database include:
  - Updates to the data stored in the database are achieved with a minimal number of operations thus reducing the opportunities for data inconsistencies.
  - Reduction in the file storage space required by the base relations thus minimizing costs.

Data Redundancy and Update Anomalies

- Problems associated with data redundancy are illustrated by comparing the Staff and Branch relations with the Staff-Branch relation.

Data Redundancy and Update Anomalies

- Problems associated with data redundancy are illustrated by comparing the Staff and Branch relations with the Staff-Branch relation.

Data Redundancy and Update Anomalies

Staff

staffNo	sName	position	salary	branchNo
SL21	John White	Manager	30000	B005
SG37	Ann Beech	Assistant	12000	B003
SG14	David Ford	Supervisor	18000	B003
SA9	Mary Howe	Assistant	9000	B007
SG5	Susan Brand	Manager	24000	B003
SL41	Julie Lee	Assistant	9000	B005

Branch

branchNo	bAddress
B005	22 Deer Rd, London
B007	16 Argyll St, Aberdeen
B003	163 Main St, Glasgow

Staff Branch

staffNo	sName	position	salary	branchNo	bAddress
SL21	John White	Manager	30000	B005	22 Deer Rd, London
SG37	Ann Beech	Assistant	12000	B003	163 Main St, Glasgow
SG14	David Ford	Supervisor	18000	B003	163 Main St, Glasgow
SA9	Mary Howe	Assistant	9000	B007	16 Argyll St, Aberdeen
SG5	Susan Brand	Manager	24000	B003	163 Main St, Glasgow
SL41	Julie Lee	Assistant	9000	B005	22 Deer Rd, London

Data Redundancy and Update Anomalies

- Staff-Branch relation has redundant data; the details of a branch are repeated for every member of staff.
- In contrast, the branch information appears only once for each branch in the Branch relation and only the branch number (branchNo) is repeated in the Staff relation, to represent where each member of staff is located.

Data Redundancy and Update Anomalies

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

Our Example

- The DreamHome Customer Rental Details form holds details about property rented by a given customer.
  - To simplify things, we will assume that a renter rents a given property once and only one property at a time.

Our Original Table

CustNo	Cname	PropNo	PAddr	RntSt	RntFnsh	Rent	OwnerNo	OName
CR76	John Kay	PG4	6 Lawrence St, Elmont	7/1/10	8/31/06	700	CO40	Tina Murphy
		PG16	5 Nova Dr, East Meadow	9/1/06	9/1/08	900	CO93	Tony Shaw
CR56	Aline Stewart	PG4	6 Lawrence St, Elmont	9/1/02	6/10/04	700	CO40	Tina Murphy
		PG36	2 Manor Rd Scarsdale	8/1/04	12/1/05	750	CO93	Tony Shaw
		PG16	5 Nova Dr, East Meadow	8/1/06	9/1/10	900	CO93	Tony Shaw

First Normal Form (1NF)

- Unnormalized – A table with one or more repeating groups.
- First Normal Form (1NF) – A relation in which the intersection of each row and column contains one and only one value

Repeating Groups

- Any collection of attributes that repeat provides a complication for a database, both in terms of storing it (how many repeating groups would you allow for) as well as querying them.
- It is necessary to recognize them so we can eliminate them.
- E.g.,

  Repeating Group = (Property_no, Paddress, RentStart, RentFinish, Rent, Owner_No, OName)

Our Table in 1NF

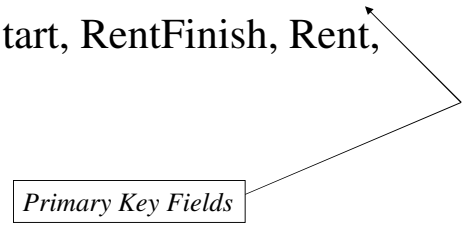
CustNo	CName	PropNo	PAddr	RntSt	RntFnsh	Rent	OwnerNo	OName
CR76	John Kay	PG4	6 Lawrence St, Elmont	7/1/10	8/31/06	700	CO40	Tina Murphy
CR76	John Kay	PG16	5 Nova Dr, East Meadow	9/1/06	9/1/08	900	CO93	Tony Shaw
CR56	Aline Stewart	PG4	6 Lawrence St, Elmont	9/1/02	6/10/04	700	CO40	Tina Murphy
CR56	Aline Stewart	PG36	2 Manor Rd Scarsdale	1/1/04	12/1/05	750	CO93	Tony Shaw
CR56	Aline Stewart	PG16	5 Nova Dr, East Meadow	8/1/06	9/1/08	900	CO93	Tony Shaw

Candidate Keys

- A candidate key for a given table is
  - unique (only one row has that value or combination of values)
  - irreducible (there is no subset of the candidate that is unique).
- Our candidate keys are:
  - (Customer_No, Property_No)
  - (Customer_No, RentStart)
  - (Property_No, RentStart)

The Customer_Rental Relation

Customer_Rental(Customer_No, Property_No, Cname, Paddress, RentStart, RentFinish, Rent, Owner_No, Oname)



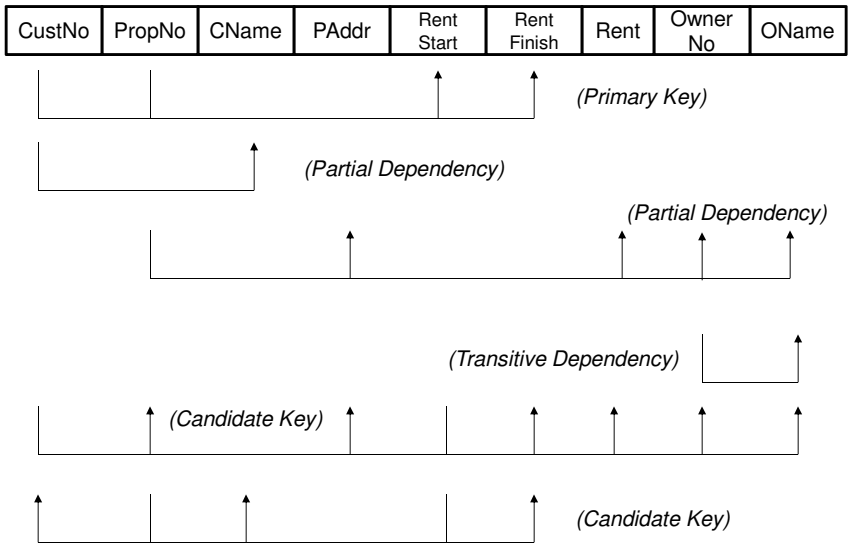
Functional Dependency

- If A and B are attributes of Relation R, B is functionally dependent on A (A→B) if each value of A is associated with one and only one value of B.
- B is fully functionally dependent on A if B is functional dependent on A and not on a proper subset of A.
- B is partially functionally dependent on A if there is some attribute that can be removed from A and the dependence still holds.

Listing All The Functional Dependencies

1. Cust_No, Prop_no → RentStart, RentFinish *(Primary Key)*
2. CustNo → Cname (Partial Dependency)
3. Prop_no → Paddress, Rent, Owner_No, Oname *(Partial Dependency)*
4. Owner_No → Oname *(Transitive Dependency)*
5. CustNo, RentStart → PropNo, Paddress, RentFinish, Rent, Owner_No, Oname *(Candidate Key)*
6. Prop_No, RentStart → CustNo, Cname, RentFinish *(Candidate Key)*

Functional Dependencies in Graphical Form



Functional Dependency in Our Table

- We have three relations with the following functional dependencies:
  - CustNo, PropNo → RentStart, RentFinish
  - CustNo → CustName
  - PropNo → Paddress, Rent, OwnerName, Oname
- Therefore, we have:
  - Customer(CustNo, Cname)
  - Rental(CustNo, PropNo, RentStart, RentFinish)
  - Property_Owner(PropNo, Paddress, Rent, OwnerNo, OName)

Second Normal Form (2NF)

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

Our Database in 2NF

Customer Relation

CustNo	CName
CR76	John Kay
CR56	Aline Stewart

Our Database in 2NF

Rentals Relation

CustNo	PropNo	RentStart	RentFinish
CR76	PG4	7/1/10	8/31/06
CR76	PG16	9/1/06	9/1/08
CR56	PG4	9/1/02	6/10/04
CR56	PG36	1/1/04	12/1/05
CR56	PG16	8/1/06	9/1/10

Our Database in 2NF

Property-Owner Relation

PropNo	PAddr	Rent	OwnerNo	OName
PG4	6 Lawrence St, Elmont	700	CO40	Tina Murphy
PG16	5 Nova Dr, East Meadow	900	CO93	Tony Shaw
PG4	6 Lawrence St, Elmont	700	CO40	Tina Murphy
PG36	2 Manor Rd Scarsdale	750	CO93	Tony Shaw
PG16	5 Nova Dr, East Meadow	900	CO93	Tony Shaw

Transitive Dependency

- If A, B, and C are attributes of a relation R such that if A→B and B→C, then C is transitively dependent on A via B.

Third Normal Form

- A relation is in 3NF is if it is 2NF and there are no non-primary-key attributes that are transitively dependent on the primary key.

Functional Dependencies in 2NF

- Customer
  - CustNo → Cname
- Rental
  - CustNo, PropNo → RentStart, RentFinish
  - PropNo, RentStart → CustNo, RentFinish
- PropertyOwner
  - PropNo → Paddr, Rent, OwnerNo, OName
  - OwnerNo → Oname *(Oname is not f.d. on PropNo)*

Our 3NF Relations

We have 4 relations:

- Customer(CustNo, Cname)
- Rental(CustNo, PropNo, RentStart, RentFinish)
- Property_For_Rent(PropNo, Paddress, Rent, OwnerNo)
- Owner(OwnerNo, OName)

Our Database in 3NF

Customer Relation

CustNo	CName
CR76	John Kay
CR56	Aline Stewart

Our Database in 3NF

Rentals Relation

CustNo	PropNo	RentStart	RentFinish
CR76	PG4	7/1/10	8/31/06
CR76	PG16	9/1/06	9/1/08
CR56	PG4	9/1/02	6/10/04
CR56	PG36	1/1/04	12/1/05
CR56	PG16	8/1/06	9/1/10

Our Database in 3NF

Property-for-Rent Relation

PropNo	PAddr	Rent	OwnerNo
PG4	6 Lawrence St, Elmont	700	CO40
PG16	5 Nova Dr, East Meadow	900	CO93
PG36	2 Manor Rd, Scarsdale	750	CO93

Our Database in 3NF

Owner Relation

OwnerNo	OName
CO40	Tina Murphy
CO93	Tony Shaw