DATA BASE

Normalization

- Main objective in developing a logical data model forrelational database systems is to create an accuraterepresentation of the data, its relationships, andconstraints.
- To achieve this objective, must identify a suitable setof relations.
- Four most commonly used normal forms are first (1NF),second (2NF) and third (3NF) normal forms, and Boyce.Coddnormal form (BCNF).

Based on functional dependencies among the attributes ofa relation.

- A relation can be normalized to a specific form toprevent possible occurrence of update anomalies.

Data Redundancy

- Major aim of relational database design is to groupattributes into relations to minimize data redundancy andreduce file storage space required by base relations.
- Problems associated with data redundancy are illustratedby comparing the following Staff and Branch relationswith the StaffBranch relation.

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

StaffBranch relation has redundant data: details of abranch are repeated for every member of staff. In contrast, branch information appears only once foreach branch in Branch relation and only branchNo isrepeated in Staff relation, to represent where eachmember of staff works.

Update Anomalies

Relations that contain redundant information maypotentially suffer from update anomalies.

- Types of update anomalies include:

. Insertion,
. Deletion,
. Modification.

Lossless-join and Dependency Preservation Properties

- Two important properties of decomposition:

  Lossless-join property enables us to find any instanceof original relation from corresponding instances in thesmaller relations.

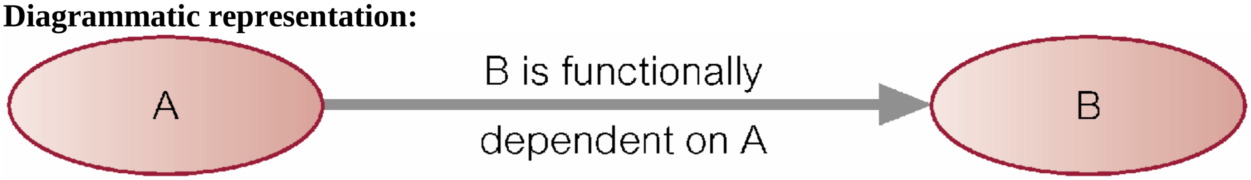
  Dependency preservation property enables us to enforcea constraint on original relation by enforcing some constrainton each of the smaller relations.

Functional Dependency

- Main concept associated with normalization.
- Functional Dependency. Describes relationship between attributes in arelation.

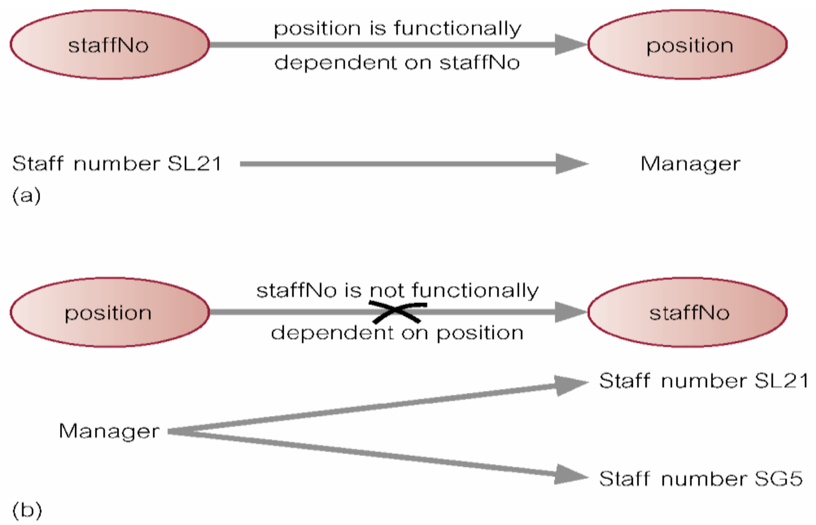
If A and B are attributes of relation R, B isfunctionally dependent on A (denoted A → B), if eachvalue of A in R is associated with exactly one valueof B in R.

Diagrammatic representation:



Determinant of a functional dependency refers toattribute or group of attributes on left-hand side of thearrow.

Example - Functional Dependency



Main characteristics of functional dependencies used innormalization:

1. **have a 1:1** relationship between attribute(s) on leftand right-hand side of a dependency; hold for all time; are nontrivial.
   - ➢ Complete set of functional dependencies for a givenrelation can be very large.
   - ➢ Important to find an approach that can reduce set to amanageable size.

➢ Need to identify set of functional dependencies (X) for arelation that is smaller than complete set of functionaldependencies (Y) for that relation and has property that every functional dependency in Y is implied by functionaldependencies in X.

➢ Set of all functional dependencies implied by a given setof functional dependencies X called closure of X (writtenX+).

- ➢ Set of inference rules, called Armstrong.s axioms,specifies how new functional dependencies can be inferredfrom given ones.
- ➢ Let A, B, and C be subsets of the attributes of relationR.

Armstrong.s axioms are as follows:

2. Reflexivity If B is a subset of A, then A ➡ B
3. Augmentation If A➡B then AC ➡BC
4. Transitivity If A➡ B and B ➡ C, then A➡C

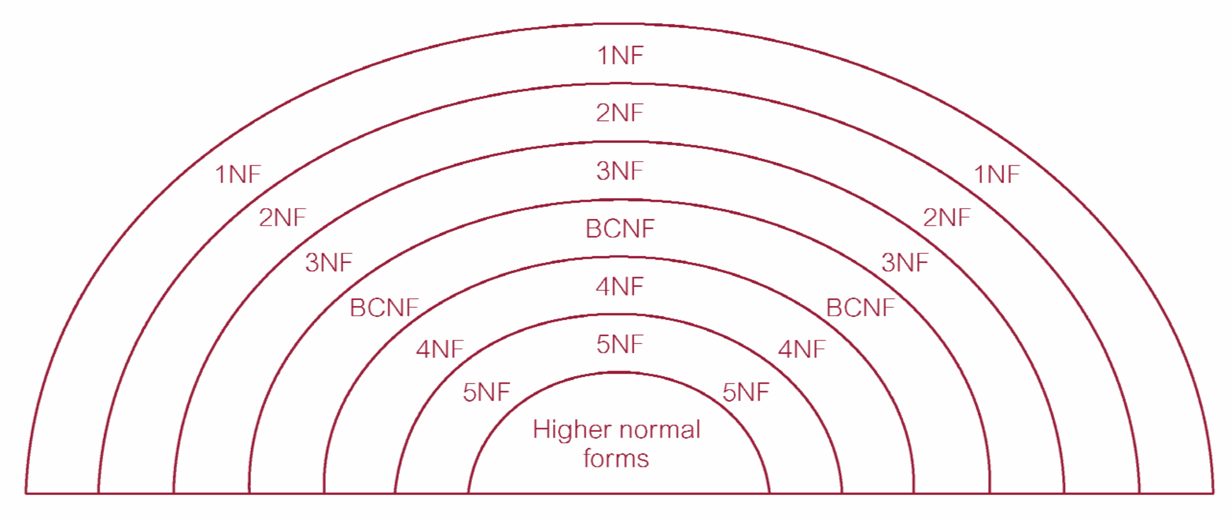
The Process of Normalization

- ➢ Formal technique for analyzing a relation based on itsprimary key and functional dependencies between itsattributes.
- ➢ Often executed as a series of steps. Each stepcorresponds to a specific normal

form, which has knownproperties.

- As normalization proceeds, relations become progressivelymore restricted (stronger) in format and also lessvulnerable to update anomalies.

Relationship Between Normal Forms



Unnormalized Form (UNF)

- A table that contains one or more repeating groups.
- To create an unnormalized table:.transform data from information source (e.g. form)into table format with columns and rows.

First Normal Form (1NF)

- A relation in which intersection of each row and columncontains one and only one value.

UNF to 1NF

- Nominate an attribute or group of attributes to act asthe key for the unnormalized table.
- Identify repeating group(s) in unnormalized table whichrepeats for the key attribute(s).
- Remove repeating group by:. entering appropriate data into the empty columns of rows containing repeating data (.flattening. thetable).

Or

By placing repeating data along with copy of theoriginal key attribute(s) into a separate relation.

Second Normal Form (2NF)

- Based on concept of full functional dependency:. A and B are attributes of a relation,. B is fully dependent on A if B is functionallydependent on A but not on any proper subset of A.
- 2NF - A relation that is in 1NF and every non-primary-keyattribute is fully

functionally dependent on the primarykey.

1NF to 2NF

- Identify primary key for the 1NF relation.
- Identify functional dependencies in the relation.
- If partial dependencies exist on the primary key removethem by placing them in a
- new relation along with copy oftheir determinant

Third Normal Form (3NF)

- Based on concept of transitive dependency:. A, B and C are attributes of a relation such that ifA B and B C,. then C is transitively dependent on A through B. (Provided that A is not functionally dependent on Bor C).
- 3NF - A relation that is in 1NF and 2NF and in which nonon-primary-key attribute is transitively dependent onthe primary key.

2NF to 3NF

- Identify the primary key in the 2NF relation.
- Identify functional dependencies in the relation.

If transitive dependencies exist on the primary keyremove them by placing them in a new relation along withcopy of their determinant.

General Definitions of 2NF and 3NF

- Second normal form (2NF): . A relation that is in 1NF and every non-primary-key attribute is fully functionally dependent on *anycandidate key*.
- Third normal form (3NF) :. A relation that is in 1NF and 2NF and in which no non-primary-key attribute is transitively dependenton *any candidate key*.

Boyce.Codd Normal Form (BCNF)

- Based on functional dependencies that take into accountall candidate keys in a relation, however BCNF also hasadditional constraints compared with general definitionof 3NF.
- BCNF - A relation is in BCNF if and only if everydeterminant is a candidate key.
- Difference between 3NF and BCNF is that for a functionaldependency A □ B, 3NF allows this dependency in arelation if B is a primary-key attribute and A is not a candidate key.
- Whereas, BCNF insists that for this dependency to remainin a relation, A must be a candidate key.
- Every relation in BCNF is also in 3NF. However, relationin 3NF may not be in BCNF.
- Violation of BCNF is quite rare.
- Potential to violate BCNF may occur in a relation that:. contains two (or more) composite candidate keys;. the candidate keys overlap (i.e. have at least one attribute in common).