The Process of Normalisation

The Process of Normalisation

- Formal technique for analysing 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.

Identifying Functional Dependencies

- Identifying all functional dependencies between a set of attributes is relatively simple if the meaning of each attribute and the relationships between the attributes are well understood.
- This information should be provided by the enterprise in the form of discussions with users and/or documentation such as the users' requirements specification.

Identifying Functional Dependencies

- However, if the users are unavailable for consultation and/or the documentation is incomplete then depending on the database application it may be necessary for the database designer to use their common sense and/or experience to provide the missing information.

Example - Identifying a set of functional dependencies for the StaffBranch relation

- Examine semantics of attributes in StaffBranch relation. Assume that position held and branch determine a member of staff's salary.

StaffBranch

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

Example - Identifying a set of functional dependencies for the StaffBranch relation

- With sufficient information available, identify the functional dependencies for the StaffBranch relation are:

staffNo → sName, position, salary, branchNo, bAddress

branchNo → bAddress

bAddress → branchNo

branchNo, position → salary

bAddress, position → salary

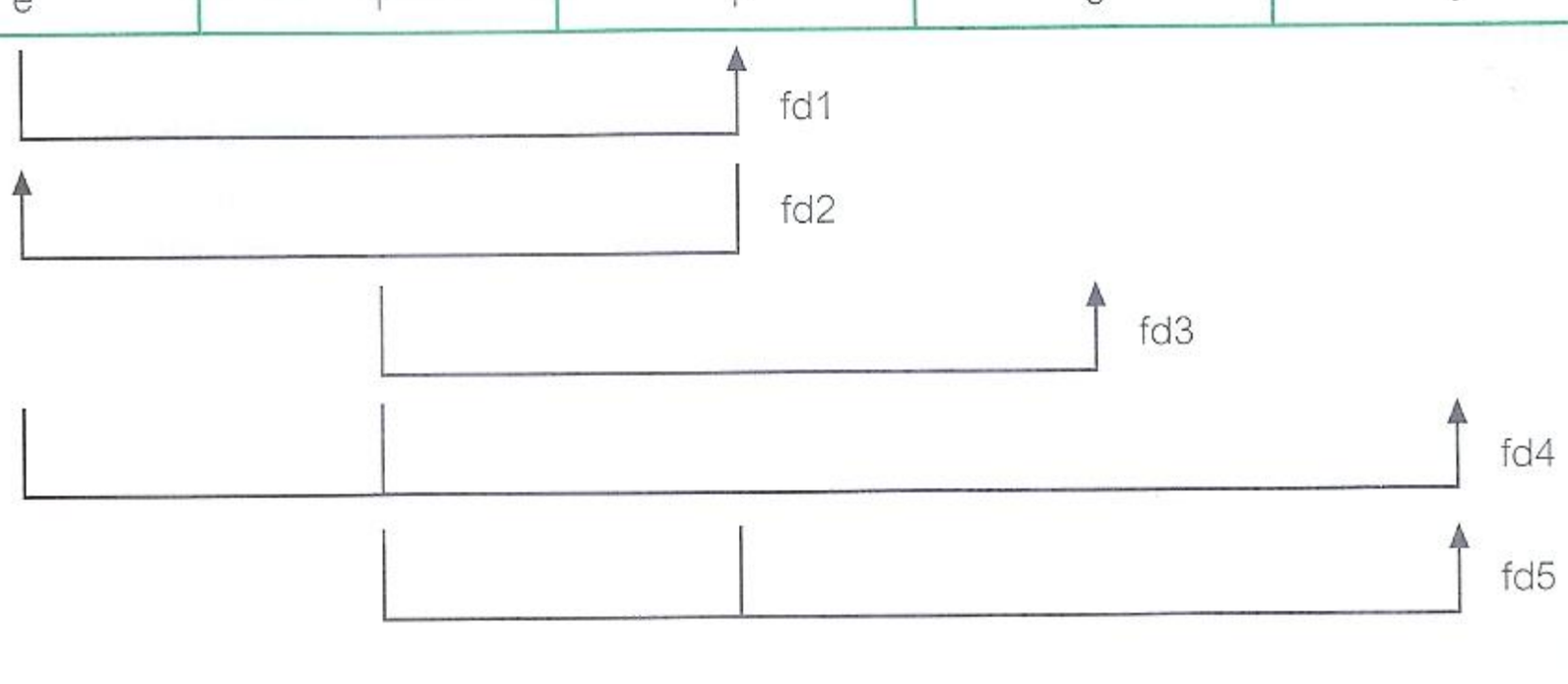
Example - Using sample data to identify functional dependencies.

- Consider the data for attributes denoted A, B, C, D, and E in the Sample relation.

- Important to establish that sample data values shown in relation are representative of all possible values that can be held by attributes A, B, C, D, and E. Assume true despite the relatively small amount of data shown in this relation.

Example - Using sample data to identify functional dependencies.

A	B	C	D	E
a	b	z	w	q
e	b	r	w	p
a	d	z	w	t
e	d	r	w	q
a	f	z	s	t
e	f	r	s	t



Example - Using sample data to identify functional dependencies

- Function dependencies between attributes A to E in the Sample relation.

  A → C (fd1)

  C → A (fd2)

  B → D (fd3)

  A, B → E (fd4)

  B, C → E (fd5)

Identifying the Primary Key for a Relation using Functional Dependencies

- Main purpose of identifying a set of functional dependencies for a relation is to specify the set of integrity constraints that must hold on a relation.

- An important integrity constraint to consider first is the identification of candidate keys, one of which is selected to be the primary key for the relation.

Example - Identify Primary Key for StaffBranch Relation

- Recall the StaffBranch relation has five functional dependencies.

staffNo → sName, position, salary, branchNo, bAddress

branchNo → bAddress

bAddress → branchNo

branchNo, position → salary

bAddress, position → salary

StaffBranch

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

Example - Identify Primary Key for StaffBranch Relation

- The determinants are staffNo, branchNo, bAddress, (branchNo, position), and (bAddress, position).

- To identify all candidate key(s), identify the attribute (or group of attributes) that uniquely identifies each tuple in this relation.

Example - Identifying Primary Key for StaffBranch Relation

- All attributes that are not part of a candidate key should be functionally dependent on the key.

- The only candidate key and therefore primary key for StaffBranch relation, is staffNo, as *all* other attributes of the relation are functionally dependent on staffNo.

Example - Identifying Primary Key for Sample Relation

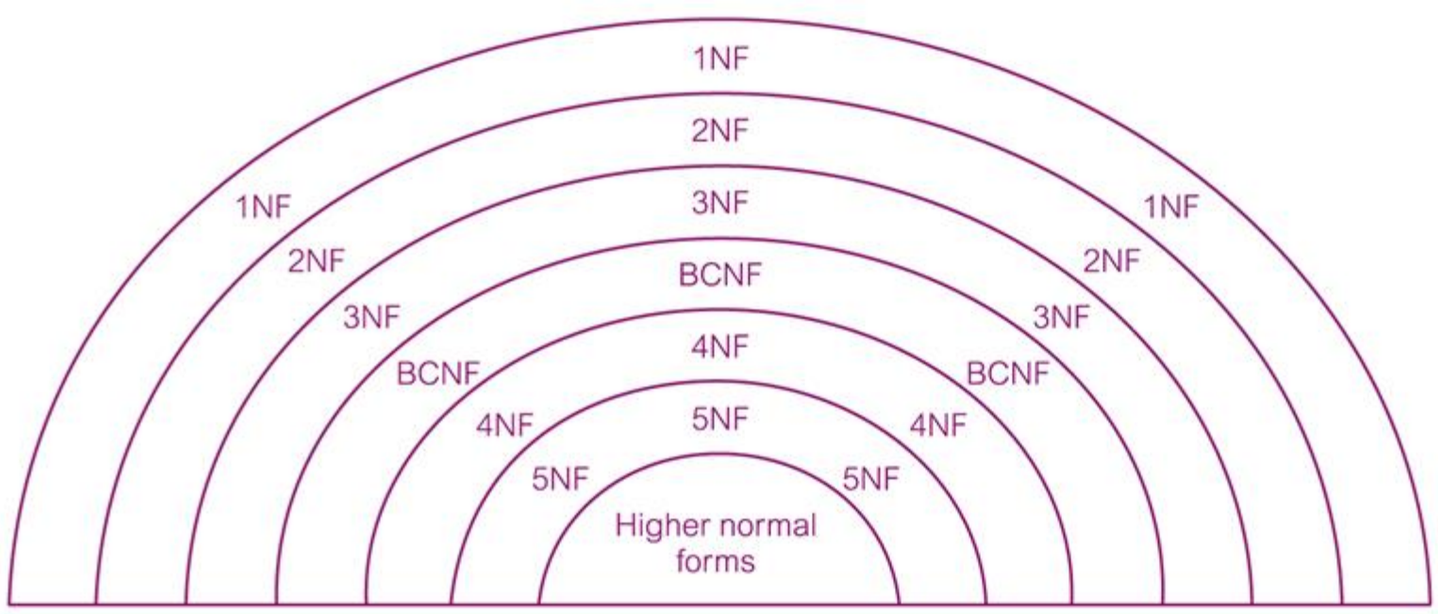
- Recall, the Sample relation has five functional dependencies. The determinants are A, B, C, (A, B), and (B, C).

- (A, B) and (B, C) determine all the other attributes of the relation.
  - In the case of (A, B), A → C, B → D, and (A, B)→ E.
  - The attributes that make up the determinant (A, B) can determine all the other attributes in the relation either separately as A or B or together as (A, B).
  - Likewise, in the case of (B, C), B → D, C → A, and (B, C) → E

- (A, B) and (B, C) are identified as candidate key values for this relation.

The Process of Normalisation

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



Diagrammatic illustration of the relationship between the normal forms.

The Process of Normalisation