

Database Creation and Management

Normalization

Sasmita Kumari Nayak
CSE Department

Normalization

- A technique for producing a set of tables with minimal redundancy that support the data requirements of an organization.

Data redundancy

- Major aim of relational database design is to group columns into tables to minimize data redundancy and reduce file storage space required by implemented base tables.
- Problems associated with data redundancy are illustrated by comparing the Staff and Branch tables with the StaffBranch table.

Staff and DistributionCenter tables with StaffDistributionCenter table

Staff

staffNo	name	position	salary	dCenterNo
S1500	Tom Daniels	Manager	48000	D001
S0003	Sally Adams	Assistant	30000	D001
S0010	Mary Martinez	Manager	51000	D002
S3250	Robert Chin	Assistant	33000	D002
S2250	Sally Stern	Manager	48000	D004
S0415	Art Peters	Manager	42000	D003

DistributionCenter

dCenterNo	dAddress	dTelNo
D001	8 Jefferson Way, Portland, OR 97201	503-555-3618
D002	City Center Plaza, Seattle, WA 98122	206-555-6756
D003	14 – 8th Avenue, New York, NY 10012	212-371-3000
D004	2 W. El Camino, San Francisco, CA 94087	822-555-3131

StaffDistributionCenter

staffNo	name	position	salary	dCenterNo	dAddress	dTelNo
S1500	Tom Daniels	Manager	48000	D001	8 Jefferson Way, Portland, OR 97201	503-555-3618
S0003	Sally Adams	Assistant	30000	D001	8 Jefferson Way, Portland, OR 97201	503-555-3618
S0010	Mary Martinez	Manager	51000	D002	City Center Plaza, Seattle, WA 98122	206-555-6756
S3250	Robert Chin	Assistant	33000	D002	City Center Plaza, Seattle, WA 98122	206-555-6756
S2250	Sally Stern	Manager	48000	D004	2 W. El Camino, San Francisco, CA 94087	822-555-3131
S0415	Art Peters	Manager	42000	D003	14 – 8th Avenue, New York, NY 10012	212-371-3000

Cont...

- StaffDistributionCenter table has redundant data; the details of a distribution center are repeated for every member of staff.
- In contrast, the details of each distribution center appears only once for each centre in the DistributionCenter table and only the distribution center number (dCenterNo) is repeated in the Staff table, to represent where each member of staff is located.

Anomalies

- Tables that contain redundant information may potentially suffer from anomalies.
- Problems that can occur in poorly planned or un-normalized databases where all the data is stored in one table (a flat-file database). This is called as **anomaly**.
- Types of update anomalies include:
 - insertion,
 - deletion,
 - Modification or update .

Example of Anomalies: Manufacturing company

emp_id	emp_name	emp_address	emp_dept
101	Rick	Delhi	D001
101	Rick	Delhi	D002
123	Maggie	Agra	D890
166	Glenn	Chennai	D900
166	Glenn	Chennai	D004

Update anomaly:

- An update anomaly exists when one or more instances of duplicated data are updated, but not all.
- From **Manufacturing company** table we have two rows for employee Rick as he belongs to two departments of the company. If we want to update the address of Rick then we have to update the same in two rows or the data will become inconsistent. If somehow, the correct address gets updated in one department but not in other then as per the database, Rick would be having two different addresses, which is not correct and would lead to inconsistent data.

Insert anomaly:

- An insert anomaly occurs when certain attributes cannot be inserted into the database without the presence of other attributes.
- Suppose a new employee joins the company, who is under training and currently not assigned to any department then we would not be able to insert the data into the table if emp_dept field doesn't allow nulls (then we have to insert NULL there, leading to insertion anomaly).

Delete anomaly:

- A delete anomaly exists when certain attributes are lost because of the deletion of other attributes.
- Suppose, if at a point of time the company closes the department D890 then deleting the rows that are having emp_dept as D890 would also delete the information of employee Maggie since she is assigned only to this department.