

Data(base) normalisation



Tamas Piros

Normalisation



- The process of removing duplicate data (data redundancy) and therefore improving data integrity
- In database 2 types of data integrity exists
 - Entity integrity: via a Primary Key
 - Referential integrity: via a foreign key
- Uses “normal forms”
 - First normal form (1NF) introduced in 1970
 - Sixth normal form (6NF) introduced in 2003

Normalisation



- Normalisation must happen in steps from UNF to 6NF (if needed)
- Each *NF defined a set of requirements
 - 1NF: no repeating groups, no duplicates, atomic column values
 - 2NF: remove duplicates
 - 3NF: each data has it's own column

Example



Name	Department	Salary	Email
John Smith	IT	45000	john.smith@corp.org
Adam Black	IT	43900	a.smith@corp.org adam.smith@corp.org
Susan White	HR	44500	susan@corp.org
Kate Brown	Finance	24500	kate@corp.org
John Smith	HR	43000	john@corp.org

Example



ID (PK)	Name	Department	Salary	Email
1	John Smith	IT	45000	john.smith@corp.org
2	Adam Black	IT	43900	a.smith@corp.org adam.smith@corp.org
3	Susan White	HR	44500	susan@corp.org
4	Kate Brown	Finance	24500	kate@corp.org
5	John Smith	HR	43000	john@corp.org

Example



ID (PK)	Name	Department (FK)	Salary	Email
1	John Smith	1	45000	john.smith@corp.org
2	Adam Black	1	43900	a.smith@corp.org adam.smith@corp.org
3	Susan White	2	44500	susan@corp.org
4	Kate Brown	3	24500	kate@corp.org
5	John Smith	2	43000	john@corp.org

ID	Name
1	IT
2	HR
3	Finance

Example



ID (PK)	Name	Department (FK)	Salary	Email1	Email2
1	John Smith	1	45000	john.smith@corp.org	NULL
2	Adam Black	1	43900	a.smith@corp.org	adam.smith@corp.org
3	Susan White	2	44500	susan@corp.org	NULL
4	Kate Brown	3	24500	kate@corp.org	NULL
5	John Smith	2	43000	john@corp.org	NULL

ID	Name
1	IT
2	HR
3	Finance