Information
Technology
Institute

Normalization

Lecturer: Manar Ayman
Mail: mafathi@iti.gov.eg
Room: 3005
Made by: Shahinaz S. Azab
Edited by: Mona Saleh , Rana Salah, Manar Ayman

What is Normalization?

- **Normalization of data:** a process that takes a table through a series of tests (normal forms) to certify the goodness of a design and thus to minimize redundancy and anomalies (insert, update, delete anomalies)

Normalization Approaches

- **A bottom-up design methodology** considers the basic relationships among individual attributes as the starting point and uses those to construct relation schemas (tables). This approach is not very popular in practice.

- **A top-down design methodology** Perform a conceptual schema design using a conceptual model such as ER and map the conceptual design into a set of relations.

Why do we need Normalization?

Redundancy

EMP_DEPT

Ename	Ssn	Bdate	Address	Dnumber	Dname	Dmgr_ssn
Smith, John B.	123456789	1965-01-09	731 Fondren, Houston, TX	5	Research	333445555
Wong, Franklin T.	333445555	1955-12-08	638 Voss, Houston, TX	5	Research	333445555
Zelaya, Alicia J.	999887777	1968-07-19	3321 Castle, Spring, TX	4	Administration	987654321
Wallace, Jennifer S.	987654321	1941-06-20	291 Berry, Bellaire, TX	4	Administration	987654321
Narayan, Ramesh K.	666884444	1962-09-15	975 FireOak, Humble, TX	5	Research	333445555
English, Joyce A.	453453453	1972-07-31	5631 Rice, Houston, TX	5	Research	333445555
Jabbar, Ahmad V.	987987987	1969-03-29	980 Dallas, Houston, TX	4	Administration	987654321
Borg, James E.	888665555	1937-11-10	450 Stone, Houston, TX	1	Headquarters	888665555

Redundancy Redundancy

EMP_PROJ

Ssn	Pnumber	Hours	Ename	Pname	Plocation
123456789	1	32.5	Smith, John B.	ProductX	Bellaire
123456789	2	7.5	Smith, John B.	ProductY	Sugarland
666884444	3	40.0	Narayan, Ramesh K.	ProductZ	Houston
453453453	1	20.0	English, Joyce A.	ProductX	Bellaire
453453453	2	20.0	English, Joyce A.	ProductY	Sugarland
333445555	2	10.0	Wong, Franklin T.	ProductY	Sugarland
333445555	3	10.0	Wong, Franklin T.	ProductZ	Houston
333445555	10	10.0	Wong, Franklin T.	Computerization	Stafford
333445555	20	10.0	Wong, Franklin T.	Reorganization	Houston
999887777	30	30.0	Zelaya, Alicia J.	Newbenefits	Stafford
999887777	10	10.0	Zelaya, Alicia J.	Computerization	Stafford
987987987	10	35.0	Jabbar, Ahmad V.	Computerization	Stafford
987987987	30	5.0	Jabbar, Ahmad V.	Newbenefits	Stafford
987654321	30	20.0	Wallace, Jennifer S.	Newbenefits	Stafford
987654321	20	15.0	Wallace, Jennifer S.	Reorganization	Houston
888665555	20	Null	Borg, James E.	Reorganization	Houston

Normalization Avoids

- Duplication of Data
- Insert Anomaly
- Delete Anomaly
- Update Anomaly
- Frequent Null Values

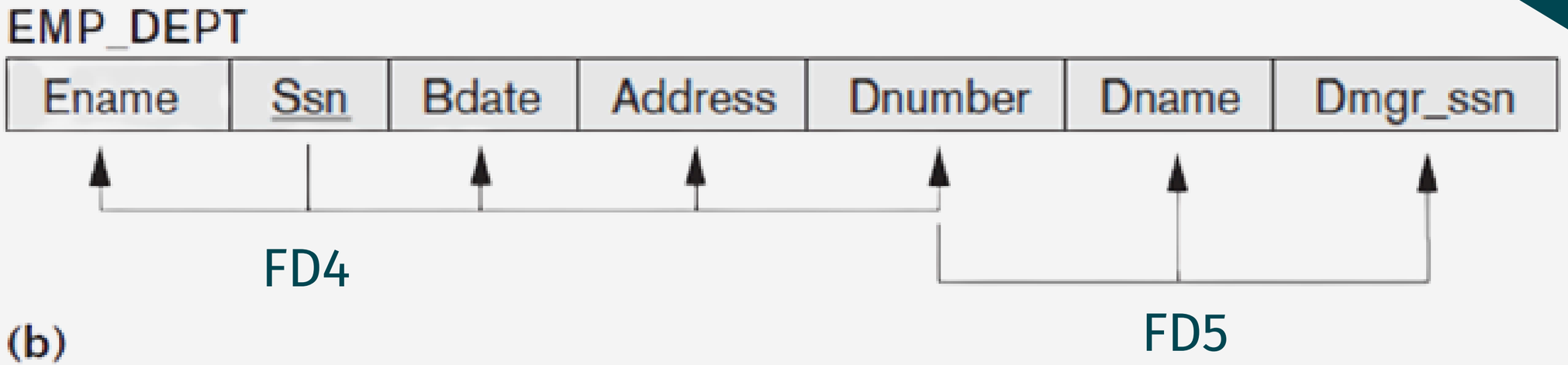
When to use Normalization?

- To Certify the goodness of relational schema design.
- When acquiring existing database design from previous legacy models, or from existing files.

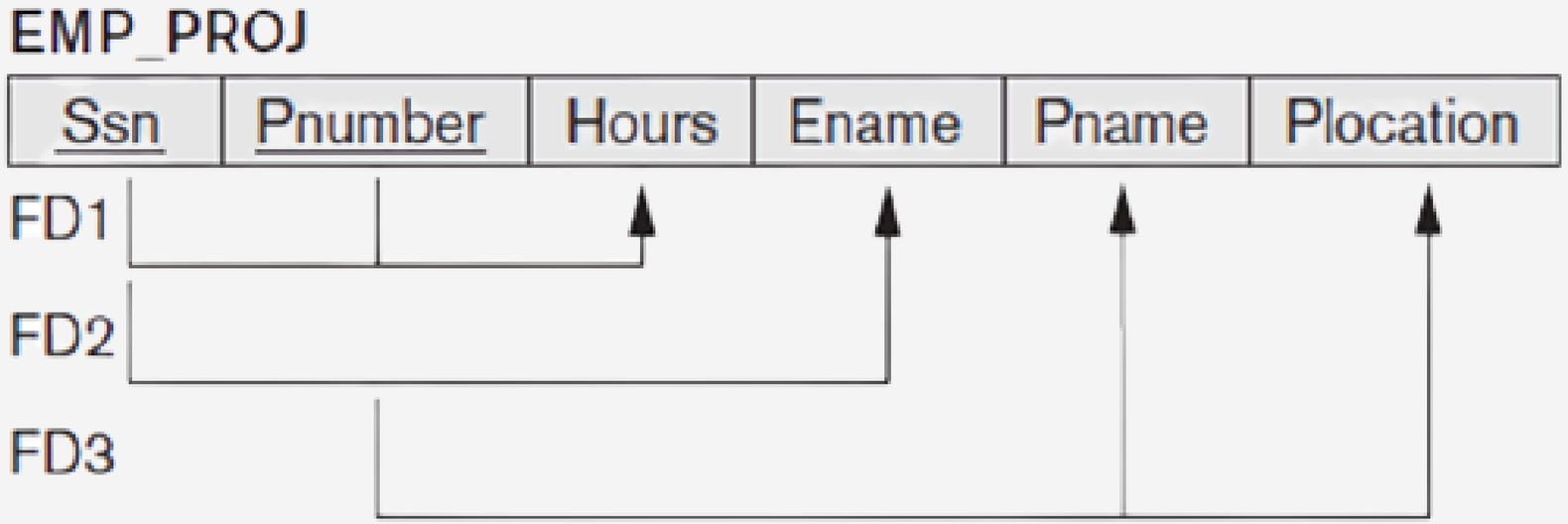
Functional Dependency

- A Constriant between 2 attributes (columns) or two sets of columns.
- A -> B if “for every valid instance of A, the value of A uniquely determines the value of B”.
- Or …… A -> B if “there exists at most one value of B for every value of A”.

Types of Functional Dependency



(b)



Types of Functional Dependency

- Full Functional Dependency- X → Y is a FFD if removal of any attribute A from X means that the dependency does not hold any more.

- Partial Functional Dependency- X→Y is a PFD if some attribute A ε X can be removed from X and the dependency still holds.

- Transitive Functional Dependency- X→Y in a relation R is a TFD if there exists a set of attributes Z in R that is neither a primary key nor a subset of any key of R, and both X→Z and Z→Y hold.

Definition

- **Normalization:**
  - The process of decomposing unsatisfactory "bad" relations by breaking up their attributes into smaller relations
- **Normal Form:**
  - Condition using keys and FDs of a relation to certify whether a relation schema is in a particular normal form.

First Normal Form (1NF):

- A relation is in 1NF if it contains no multivalued, repeating groups or composite attributes.

- **To put a relation in 1NF**
- Remove each repeating group and place it in a new table carrying the PK as a FK
- Remove each multivalued attribute and place it in a new table carrying the PK as a FK
- Put composite attribute subparts each in a column when necessary.

School Example:

Stud_ID	Name	Loc	Tel	Level	Level_mgr	Subject	Subj_desc	G
11	Ali	Cairo	010...	Primary	Noha M.	DB, CN	Database, Networks	A, B
22	Mai	Giza	011..., 010...	Primary	Noha M.	CN,DB	Networks, Database	B, C
33	Marwa	Giza	010...	Secondary	Moh. A.	SW, DB	Software, Database	A, A

School Example: 1NF

Stud_ID	Name	Location	Level	Level_Mgr

Stud_ID	Tel

Stud_ID	Subject	Subject_Desc	Grade

Second Normal Form: (2NF)

- A relation is in 2NF if it is in 1NF and every nonkey attribute is not partially dependent on the primary key

- **To put a relation in 2NF:**
  - Remove partial functional dependent non-keys carrying the key they depend on and place them in a new table.

Guidelines

- A relation is in 2NF if it is in 1NF and any one of these is true:
  - The PK consists of only 1 attribute.
  - All attributes are part of the PK (no nonkey attributes).
  - Every non key attribute is functionally dependent on the whole PK.

School Example: 2NF

Stud_ID	Name	Location	Level	Level_Mgr

Stud_ID	Tel

Stud_ID	Subject	Grade

Stud_ID	Subject_Desc

Third Normal Form: (3NF)

- A relation is in **3NF** if it is in 2NF and no transitive dependencies exist

- **To put a relation in 3NF:**
  - Remove the nonkey attributes carrying the nonkey attribute they depend on and place them in a new table. (Hint: leave the nonkey they depend on in the same table as well).

School Example 3NF

Student	Stud_ID	Name	Location	Level
Student_Tel	Stud_ID	Tel		
Level	Level	Level_Mgr		
Stud_Subject	Stud_ID	Subject	Grade	
Subject	Stud_ID	Subject_Desc		

ITI Example

ITI Students Sheet

Student Number: ITI205-40
Student Name: Hassan Ali Ahmed
Address (Street, City): 12 Haram st, giza
Tel no/Mobile: 33868420, 01111111253

F-code: ENG
Faculty: Engineering
Major: Computer

Department Name	Department Description	Admission grade	Comments
ERP - SAP	ERP-SAP Functional Consultant	59	Average Personality
Java - MAD	Java mobile applications developer	70	Very Good
CS	Cyber Security	60	Above avgerage technical

ITI Example: 1NF

Student

(Stud_No,Stud_Name, F-code, Faculty, Major, Street, City)

Student_Tel

(Stud_No, Tel_No)

Department_Student

(Dept_Name, Stud_No, Dept_desc , Ad_Grade, Comments)

ITI Example: 2NF

Student

(Stud_No,Stud_Name, F-code, Faculty, Major, Street, City)

Student_Tel

(Stud_No, Tel_No)

Department_Student

(Dept_Name, Stud_No , Ad_Grade, Comments)

Department

(Dept_name, Dept_Desc)

ITI Example: 3NF

Student

(Stud_No,Stud_Name, F-code, Major, Street, City)

Faculty

(F-code, Faculty)

Student_Tel

(Stud_No, Tel_No)

Department_Student

(Dept_Name, Stud_No , Ad_Grade, Comments)

Department

(Dept_name, Dept_Desc)

Real World - School Data

Student First	Parent 1	Parent 2	Application No	
Renee	Ann Jones	Theodore Smith	123	
Lucy	Barbara Mills	Steve Mills	558	
Brendan	Jennifer Jones	Stephen Jones	145	…
City	Postal Code	Birth date	Previous Teacher	Current Teacher
Annandale	22003	6/25/1983	Hamil	Burke
Annandale	22003	8/14/1983	Hamil	Burke
Fairfax	22032	6/13/1984	Hamil	Burke
Student_Phone	Course	Course-desc	Enrolled	Attended/ days
(703) 323-0893, (703) 3240708	X,Y,Z	X,y,z	96/97, 96/97, 97/98	0,0,0
(703) 764-5829	Y	Y	96/97	0
(703) 978-1083	Z	Z	96/97	0

ONF

- **Student**
  - (App_No, Stud_Fname, Parent1, Parent2, City, Postal_Code, Birthdate, Prev_Teacher, Curr_Teacher, Student_Phone, Course, Course_Desc, Enrolled, Att_Days)

1NF

- **Student**
  - (App_No, Stud_Fname, Parent1, Parent2, City, Postal_Code, Birthdate, Prev_Teacher, Curr_Teacher)

- **Student_Course**
  - (App_No,Course, Course_Desc, Enrolled, Att_Days)

- **Student_Phone**
  - (App_No,Phone)

2NF

- **Student**
  - (App_No, Stud_Fname, Parent1, Parent2, City, Postal_Code, Birthdate, Prev_Teacher, Curr_Teacher)
- **Student_Course**
  - (App_No,Course, Enrolled, Att_Days)
- **Student_Phone**
  - (App_No,Phone)
- **Course**
  - (Course, Course_Desc)

3NF

- **Student**
  - (App_No, Stud_Fname, Parent1, Parent2, Postal_Code, Birthdate, Prev_Teacher, Curr_Teacher)
- **Student_Course**
  - (App_No,Course, Enrolled, Att_Days)
- **Student_Phone**
  - (App_No,Phone)
- **Course**
  - (Course, Course_Desc)
- **City**
  - (City, Postal_Code)

Supplier Data

- Relation (s#, country, currency, p#, qty)
- where
  - s# supplier identification number (this is the primary key)
  - country name of country where supplier is located
  - currency: Currency of the country of each supplier
  - p# part number of part supplied
  - qty quantity of parts supplied to date
- In order to uniquely associate quantity supplied (qty) with part (p#) and supplier (s#), a composite primary key composed of s# and p# is used.

Supplier Data

Sales Order

Fiction Company
202 N. Main
Mahattan, KS 66502

CustomerNumber:	*1001*	**Sales Order Number:**	405
Customer Name:	*ABC Company*	**Sales Order Date:**	2/1/2000
Customer Address:	*100 Points*	**Clerk Number:**	*210*
	Manhattan, KS 66502	**Clerk Name:**	*Martin Lawrence*

Item Ordered	Description	Quantity	Unit Price	Total
800	*widgit small*	*40*	*60.00*	*2,400.00*
801	*tingimajigger*	*20*	*20.00*	*400.00*
805	*thingibob*	*10*	*100.00*	*1,000.00*
	Order Total			*3,800.00*

1NF

- **Supplier**
  - (S#, country, currency)
- **Supplier_Parts**
  - (S#,P#,qty)

2NF

Same as First

- **Supplier**
  - (S#, country, currency)
- **Supplier_Parts**
  - (S#,P#,qty)

3NF

- **Supplier**
  - (S#, country)
- **Country**
  - (Country, Currency)
- **Supplier_Parts**
  - (S#,P#,qty)

Questions?