Cauvery College for Women (Autonomous)

Nationally Accredited (III Cycle) with 'A' Grade by NAAC
Annamalai Nagar, Tiruchirappalli-18.

Name of the Faculty : M.Ellakkiya

Designation : Assistant Professor

Department : Computer Applications

Programme :BCA

Batch : 2016-2017 Onwards

Semester : IV

Course : DataBaseSystems

Course Code :16SCCCA4

Unit : V

Topics Covered : Normal forms.

Contact No :9994683100

Database Normalization

- Normalization:

  - is the process of evaluating & correcting the structures of the tables in a database

  - The goal:

    - to minimize or remove data redundancy

    - Tooptimalize the data structure

    - Accomplished by thoroughly investigating the various data type and their relationships with one another.

- Data redundancy:

  - The repeat of key fields usages in other tables

Database Normalization

- Functional dependencies:
  - Require that the value for a certain set of attributes determines uniquely the value for another set of attributes
  - are akin to a generalization of the notion of a key
  - Let R be a relation and

    $$\alpha \subseteq R \text{ and } \beta \subseteq R$$

    The functional dependency :

    $$\alpha \rightarrow \beta$$

    holds on R and only if d or any tuples t_1 & t_2 that agree on the attributes α, they also agree on the attributes β.
  - That is, $t_1[\alpha] = t_2[\alpha] \rightarrow t_1[\beta] = t_2[\beta]$

Database Normalization

- Functional dependencies

  Example: consider student(Sid, Sname, DeptId)

  instance of student.

Sid	Sname	DeptId
CL12001	JOHN	13
CL13050	WENPENG	13
DE10003	ALDI	15
PS11123	ILJA	11
IT09256	LISANDRO	09
CL13075	MATTHEW	13

Is this true?	Yes	No
Sid → Sname		
Sid → DeptId		
Sname → DeptId		
Sname → Sid		
DeptId → Sname		
DeptId → Sid		

Database Normalization

- Functional dependencies

  Example: consider student(Sid, Sname, DeptId)

  instance of student.

Sid	Sname	DeptId
CL12001	JOHN	13
CL13050	WENPENG	13
DE10003	ALDI	15
PS11123	ILJA	11
IT09256	LISANDRO	09
CL13075	MATTHEW	13

Is this true?	Yes	No
Sid → Sname	✔	
Sid → DeptId	✔	
Sname → DeptId		✔
Sname → Sid		✔
DeptId → Sname		✔
DeptId → Sid		✔

Database Normalization

- examine the following poor database design:

	Sid	Cname	time	room	Lid
Edit 3⌐ Copy ⊖ Delete	CL0001	Machine Learning	Wed 10.15	L155	PR145
Edit 3⌐ Copy ⊖ Delete	CL0002	Information Retrieval	Tue 12.15	C131	PD220
Edit 3⌐ Copy ⊖ Delete	CL0003	Machine Learning	Wed 10.15	L155	PR145
Edit 3⌐ Copy ⊖ Delete	CL0004	Information Extraction	Thu 10.00	C149	PR111

- Problems:
    - No need to repeatedly store the class time & Professor ID
    - Which one is the key?

Database Normalization

- First Normal Form (1NF):

  - A row of data cannot contain a repeating group of data.

  - Each row of data must have a unique identifier, i.e primary key

- This can be done by

  - Eliminating the repeated groups of data through creating separate tables of related data

  - Identify each set of related data with a primary key

  - All attributes are single valued (1 data type) & non-repeating

    - Student information:

      Sid   *Sname*   *Major*   *Minor*   *IntakeYear*

    - Course information

      Cid   *Cname*   *Lid*   *Time*   *Room*

    - Lecturer Information

      Lid   *Lname*   *Ltitle*

Database Normalization

- Second Normal form (2NF):

  - A table should meet 1NF

  - There must not be any partial dependency of any column on primary key (Records should not depend on anything other than a table's primary key)

- Recall our poor database design:

  Sid → Cname or Cname → time ?

← T →				Sid	Cname	time	room	Lid
	Edit	Copy	Delete	CL0001	Machine Learning	Wed 10.15	L155	PR145
	Edit	Copy	Delete	CL0002	Information Retrieval	Tue 12.15	C131	PD220
	Edit	Copy	Delete	CL0003	Machine Learning	Wed 10.15	L155	PR145
	Edit	Copy	Delete	CL0004	Information Extraction	Thu 10.00	C149	PR111

Database Normalization

- Second Normal Form (2NF) solution:

    - **Create** separate tables for sets of values that apply to multiple records

    - **Relates** the tables with a **foreign key**

    - **Remove** subsets of data that apply to multiple rows of a table and **place** them in separate tables

      enrolled

      | Sid | Cid | grade (?) |

    - What do we do with the attribute time, room, & Lid?

Database Normalization

- Third Normal Form (3NF):

  - Eliminate all attributes (columns) that do not directly dependent upon the primary key

  - Each non-primary key attribute must be dependent only on primary key (no transitive dependency)

  - Example:

    Student:

    Sid   Sname   Major   Minor   Intake Year

    - *Which attribute is not directly dependent on Sid?*
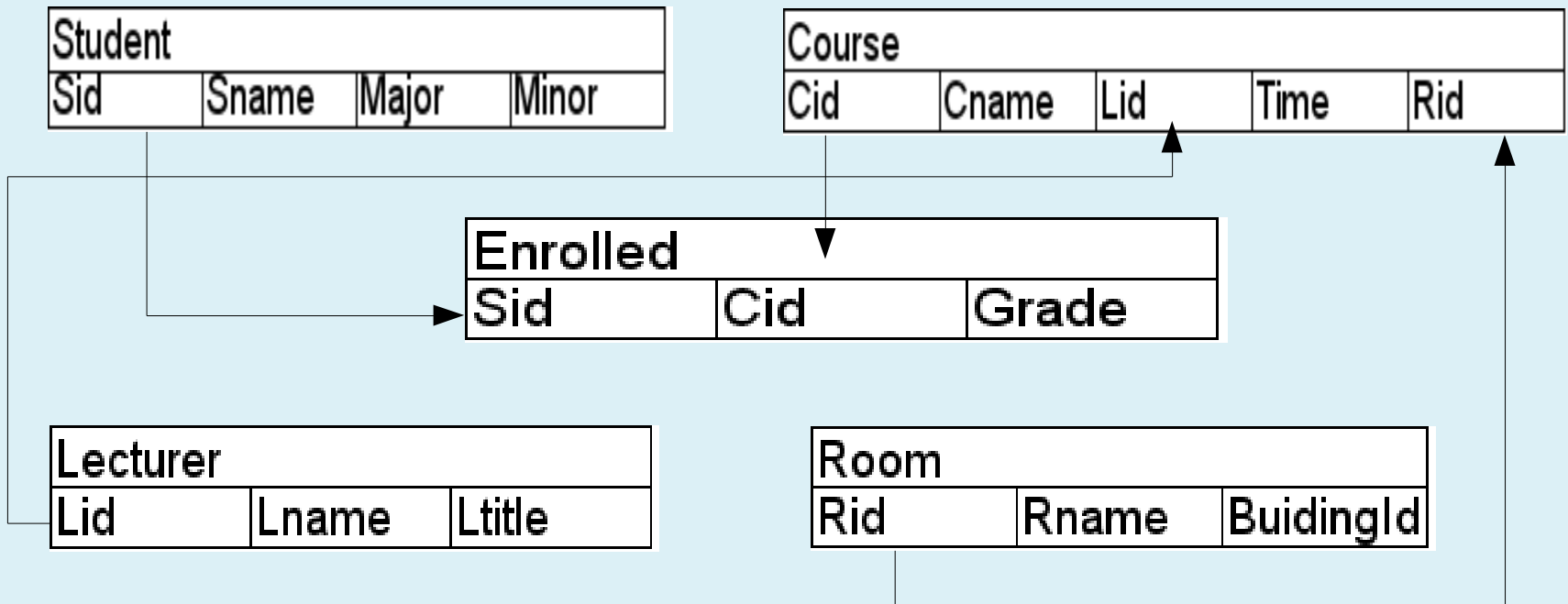
    Student:

    Sid   Sname   Major   Minor
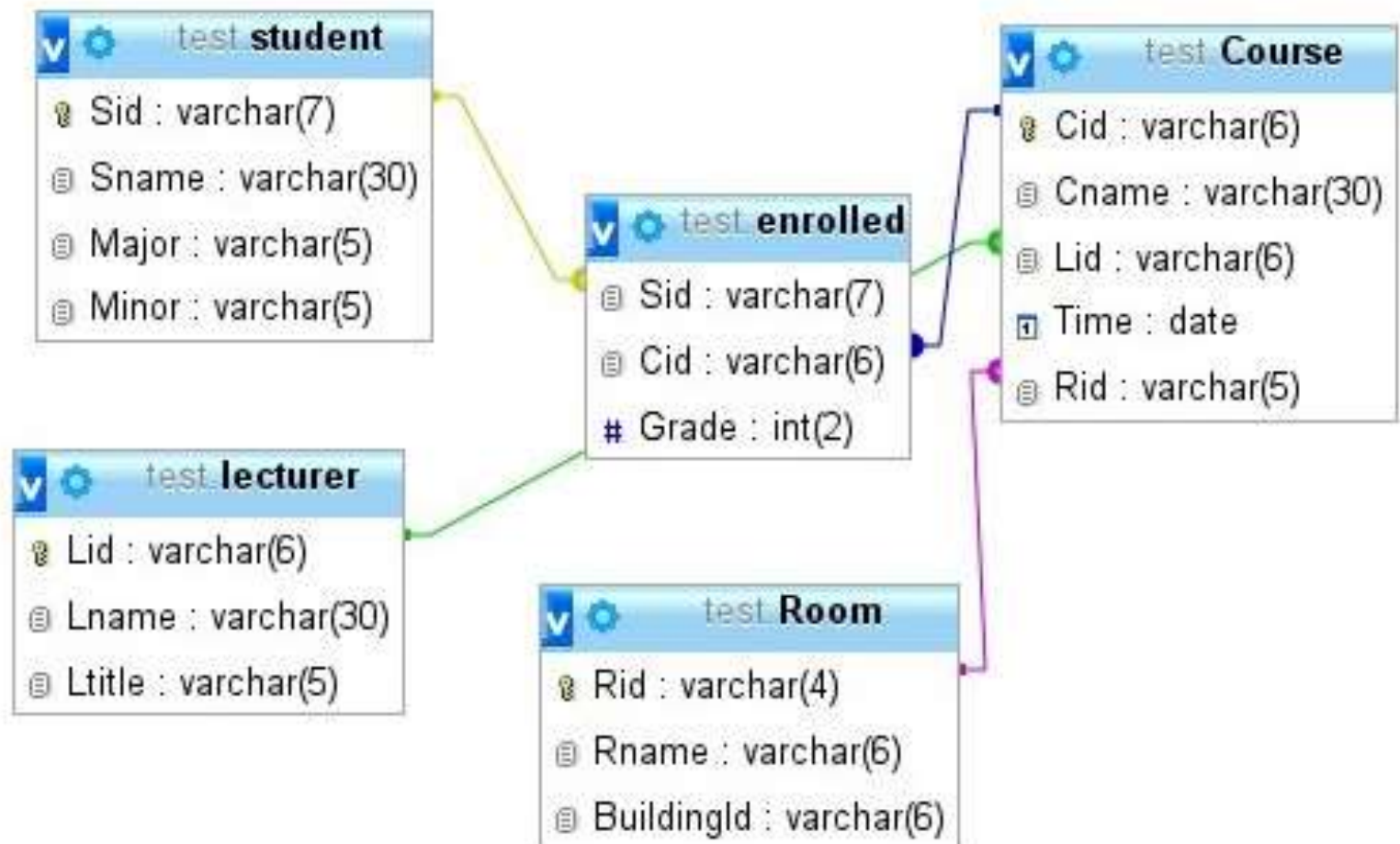
Database Normalization

- Old design

	Sid	Cname	time	room	Lid
✏ Edit ⃞ Copy ⊖ Delete	CL0001	Machine Learning	Wed 10.15	L155	PR145
✏ Edit ⃞ Copy ⊖ Delete	CL0002	Information Retrieval	Tue 12.15	C131	PD220
✏ Edit ⃞ Copy ⊖ Delete	CL0003	Machine Learning	Wed 10.15	L155	PR145
✏ Edit ⃞ Copy ⊖ Delete	CL0004	Information Extraction	Thu 10.00	C149	PR111

- New design

Student

Sid	Sname	Major	Minor

Course

Cid	Cname	Lid	Time	Rid

Enrolled

Sid	Cid	Grade

Lecturer

Lid	Lname	Ltitle

Room

Rid	Rname	BuidingId

Database Normalization

- Storing the relation among tables in database

Database Normalization

- Exercise:
  - Which normal form does this table violate?
  - And how do you normalize it?

Person	Title	Author	Pages	Year
Yakup	Database Management System	Ramakhrisnan, Raghu	903	2010
Wenpeng	Beyond Human-Computer Interaction	Preece, Jennifer	889	2009
Amy	Support Your Local Wizard	Duane, Diane	473	1990
Dwika	The Hobbit	Tolkien, JRR	389	1995
Yadoll	Beyond Human-Computer Interaction	Preece, Jennifer	889	2009
Irina	Support Your Local Wizard	Duane, Diane	473	1990