

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

شروع اللہ کے پاک نام سے جو بڑا مہربان نہایت رحم والا ہے



Dr. Abid Sohail Bhutta

abidbhutta@cuilahore.edu.pk

**Department of Computer
Science,**

**COMSATS University , Lahore
Campus**

Database Systems



Lecture 20

Schema Refinements , Normalization and 1st NF



Today's Lecture

- **Normalization**
 - **1st Normal Form**

Recall Lecture 19

Database System Project

Normalization

- What normalization is and what role it plays in database design
- About the normal forms 1NF, 2NF, 3NF, BCNF, and 4NF
- How normal forms can be transformed from lower normal forms to higher normal forms
- That normalization and E-R modeling are used concurrently to produce a good database design
- That some situations require de-normalization to generate information efficiently

What is Normalization?

- Normalization is a database design technique that reduces data redundancy and eliminates undesirable characteristics like Insertion, Update and Deletion Anomalies.
 - Normalization rules divides larger tables into smaller tables and links them using relationships.
 - The purpose of Normalization in SQL is to eliminate redundant (repetitive) data and ensure data is stored logically.
-

Database Tables and Normalization

- Table is basic building block in database design
- Table's structure is of great interest
- Two cases:
 - possible poor table structures in good database design
 - Modify existing database with existing poor table structure
- Normalization can help recognize a poor table and convert to good tables with good structure

Database Tables and Normalization

- Normalization is process for assigning attributes to entities
 - ❑ Reduces data redundancies
 - ❑ Expending entities
 - ❑ Helps eliminate data anomalies
 - ❑ Produces controlled redundancies to link tables
 - ❑ Cost more processing efforts
 - ❑ Series steps called normal forms

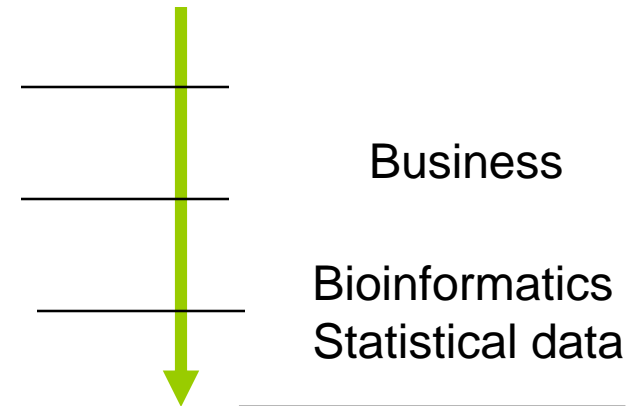
Database Tables and Normalization

■ Normalization stages

- ❑ 1NF - First normal form
- ❑ 2NF - Second normal form
- ❑ **3NF - Third normal form**
- ❑ 4NF - Fourth normal form

**Better in
dependency**

**Worse in
performance (I/O)**



Database Tables and Normalization

- Example: construction company
 - Building projects
 - Project number
 - Project name
 - Employees assigned
 - ...
 - Employee
 - Employee number
 - Employee name
 - Job classification

	PROJ NUM	PROJ NAME	EMP NUM	EMP NAME	JOB CLASS	CHG HOUR	HOURS
►	15	Evergreen	103	June E. Arbough	Elect. Engineer	\$84.50	23.8
			101	John G. News	Database Designer	\$105.00	19.4
			105	Alice K. Johnson *	Database Designer	\$105.00	35.7
			106	William Smithfield	Programmer	\$35.75	12.6
			102	David H. Senior	Systems Analyst	\$96.75	23.8
	18	Amber Wave	114	Annelise Jones	Applications Designer	\$48.10	24.6
			118	James J. Frommer	General Support	\$18.36	45.3
			104	Anne K. Ramoras *	Systems Analyst	\$96.75	32.4
			112	Darlene M. Smithson	DSS Analyst	\$45.95	44.0
	22	Rolling Tide	105	Alice K. Johnson	Database Designer	\$105.00	64.7
			104	Anne K. Ramoras	Systems Analyst	\$96.75	48.4
			113	Delbert K. Joenbrood *	Applications Designer	\$48.10	23.6
			111	Geoff B. Wabash	Clerical Support	\$26.87	22.0
			106	William Smithfield	Programmer	\$35.75	12.8
	25	Starflight	107	Maria D. Alonzo	Programmer	\$35.75	24.6
			115	Travis B. Bawangi	Systems Analyst	\$96.75	45.8
			101	John G. News *	Database Designer	\$105.00	56.3
			114	Annelise Jones	Applications Designer	\$48.10	33.1
			108	Ralph B. Washington	Systems Analyst	\$96.75	23.6
			118	James J. Frommer	General Support	\$18.36	30.5
			112	Darlene M. Smithson	DSS Analyst	\$45.95	41.4

FIGURE 4.1 A TABLE WHOSE STRUCTURE MATCHES THE REPORT FORMAT

Figure 4.1 Observations

- PRO_NUM intended to be primary key, but it contains null values.
- Table entries invite data inconsistencies

Figure 4.1 Observations

- Table displays data redundancies which yield the following anomalies
 - Update
 - Modifying JOB_CLASS
 - Insertion
 - New employee must be assigned project (phantom project)
 - Deletion
 - If employee deleted, other vital data lost

	PROJ_NUM	PROJ_NAME	EMP_NUM	EMP_NAME	JOB_CLASS	CHG_HOUR	HOURS
▶	5	Evergreen	103	June E. Arbough	Elect. Engineer	\$84.50	23.8
			101	John G. News	Database Designer	\$105.00	19.4
			105	Alice K. Johnson *	Database Designer	\$105.00	35.7
			106	William Smithfield	Programmer	\$35.75	12.6
			102	David H. Senior	Systems Analyst	\$96.75	23.8

FIGURE 4.2 THE EVERGREEN DATA

Repeating group (any project can have a group of data entries) which should **not** to be appeared in relational table

Data Organization: 1NF



	PROJ_NUM	PROJ_NAME	EMP_NUM	EMP_NAME	JOB_CLASS	CHG_HOUR	HOURS
▶	15	Evergreen	103	June E. Arbough	Elect. Engineer	\$84.50	23.8
	15	Evergreen	101	John G. News	Database Designer	\$105.00	19.4
	15	Evergreen	105	Alice K. Johnson *	Database Designer	\$105.00	35.7
	15	Evergreen	106	William Smithfield	Programmer	\$35.75	12.5
	15	Evergreen	102	David H. Senior	Systems Analyst	\$96.75	23.9
	18	Amber Wave	114	Annelise Jones	Applications Designer	\$48.10	24.6
	18	Amber Wave	118	James J. Frommer	General Support	\$18.36	45.3
	18	Amber Wave	104	Anne K. Ramoras *	Systems Analyst	\$96.75	32.1
	18	Amber Wave	112	Darlene M. Smithson	DSS Analyst	\$45.95	44.0
	22	Rolling Tide	105	Alice K. Johnson	Database Designer	\$105.00	64.7
	22	Rolling Tide	104	Anne K. Ramoras	Systems Analyst	\$96.75	48.9
	22	Rolling Tide	113	Delbert K. Joenbrood *	Applications Designer	\$48.10	23.6
	22	Rolling Tide	111	Geoff B. Wabash	Clerical Support	\$26.87	22.5
	22	Rolling Tide	106	William Smithfield	Programmer	\$35.75	12.1
	25	Starflight	107	Maria D. Alonzo	Programmer	\$35.75	24.7
	25	Starflight	115	Travis B. Bawangi	Systems Analyst	\$96.75	45.8
	25	Starflight	101	John G. News *	Database Designer	\$105.00	56.3
	25	Starflight	114	Annelise Jones	Applications Designer	\$48.10	33.1
	25	Starflight	108	Ralph B. Washington	Systems Analyst	\$96.75	23.9
	25	Starflight	118	James J. Frommer	General Support	\$18.36	30.2
	25	Starflight	112	Darlene M. Smithson	DSS Analyst	\$45.95	41.4

FIGURE 4.3 DATA ORGANIZATION: FIRST NORMAL FORM

Data Organization: 1NF



	PROJ_NUM	PROJ_NAME	EMP_NUM	EMP_NAME	JOB_CLASS	CHG_HOUR	HOURS
▶	15	Evergreen	103	June E. Arbough	Elect. Engineer	\$84.50	23.8
	15	Evergreen	101	John G. News	Database Designer	\$105.00	19.4
	15	Evergreen	105	Alice K. Johnson *	Database Designer	\$105.00	35.7
	15	Evergreen	106	William Smithfield	Programmer	\$35.75	12.5
	15	Evergreen	102	David H. Senior	Systems Analyst	\$96.75	23.9
	18	Amber Wave	114	Annelise Jones	Applications Designer	\$48.10	24.6
	18	Amber Wave	118	James J. Frommer	General Support	\$18.36	45.3
	18	Amber Wave	104	Anne K. Ramoras *	Systems Analyst	\$96.75	32.1
	18	Amber Wave	112	Darlene M. Smithson	DSS Analyst	\$45.95	44.0
	22	Rolling Tide	105	Alice K. Johnson	Database Designer	\$105.00	64.7
	22	Rolling Tide	104	Anne K. Ramoras	Systems Analyst	\$96.75	48.9
	22	Rolling Tide	113	Delbert K. Joenbrood *	Applications Designer	\$48.10	23.6
	22	Rolling Tide	111	Geoff B. Wabash	Clerical Support	\$26.87	22.5
	22	Rolling Tide	106	William Smithfield	Programmer	\$35.75	12.1
	25	Starflight	107	Maria D. Alonzo	Programmer	\$35.75	24.7
	25	Starflight	115	Travis B. Bawangi	Systems Analyst	\$96.75	45.8
	25	Starflight	101	John G. News *	Database Designer	\$105.00	56.3
	25	Starflight	114	Annelise Jones	Applications Designer	\$48.10	33.1
	25	Starflight	108	Ralph B. Washington	Systems Analyst	\$96.75	23.9
	25	Starflight	118	James J. Frommer	General Support	\$18.36	30.2
	25	Starflight	112	Darlene M. Smithson	DSS Analyst	\$45.95	41.4

FIGURE 4.3 DATA ORGANIZATION: FIRST NORMAL FORM

Conversion to 1NF

- Repeating groups must be eliminated
 - Proper primary key developed
 - Uniquely identifies attribute values (rows)
 - Combination of PROJ_NUM and EMP_NUM

In Next Lecture

- Normalization
 - **2nd Normal Form**

Thanks