DATABASE ESSENTIALS

Lecture 6

stephen.donald@cbe.ac.tz

DATABASE NORMALIZATION

- Normalization is a database design technique that reduces data redundancy and eliminates undesirable characteristics like insertion, update and deletion anomalies.

- It involves dividing large tables into smaller tables and defining the relationships between the smaller tables.

DATABASE NORMALIZATION…

- Consider the table below

Modules						
Module Name	*Staff No*	*Staff Name*	*Student No*	*Student*	*Ass Grade*	*Ass Type*
Relational Database Systems	234	Lee T	3468	Smith S	B3	Cwk1
Relational Database Systems	234	Lee T	3468	Smith S	B1	Cwk2
Relational Database Systems	234	Lee T	3778	Jones S	B2	Cwk1
Relational Database Systems	234	Lee T	3488	Patel P	B1	Cwk1
Relational Database Systems	234	Lee T	3488	Patel P	B3	Cwk2
Relational Database Design	234	Lee T	3468	Smith S	B2	Cwk1
Relational Database Design	234	Lee T	3468	Smith S	B3	Cwk2
Deductive Databases	345	Evans	3478	Smith J	A1	Exam

DATABASE NORMALIZATION…

- The following problems might arise from the above table
  - A deletion side effect/anomaly
  - An update side effect/anomaly
  - An insertion side effect/anomaly

DATABASE NORMALIZATION…

Deletion side effect

- A deletion anomaly occurs when removing data from the database results in unintended loss of information.

- If we wish to delete student 3478, the result is that we lose some valuable information.

- We lose information about deductive databases and its associated lecturer.

DATABASE NORMALIZATION…

Update side effect

- An update anomaly occurs when updating data in a database results in incomplete or inconsistent information.

- What if we change the lecturer of Deductive Databases to S Donald?

- If this information is stored in multiple records, you might accidentally update one record and forget to update the others, leading to inconsistencies.

DATABASE NORMALIZATION…

Insertion side effect

- An insertion anomaly occurs when it is not possible to add data to the database without adding additional, unnecessary information.

- What if we admit a new student on to a module with student-No 3898?

- We cannot enter a student record until a student has had at least one assessment.

DATABASE NORMALIZATION…

- Stages of normalization
  - First Normal Form (1NF)
  - Second Normal Form (2NF)
  - Third Normal Form (3NF)
  - BCNF (Boyce-Codd Normal Form)
  - 4NF (Fourth Normal Form)

- In most practical applications, normalization achieves its best in 3NF.

DATABASE NORMALIZATION…

- Un-normalized table refers to a table where all information is stored in one table.

- Example

FULL NAMES	PHYSICAL ADDRESS	MOVIES RENTED	SALUTATION
Janet Jones	Mbezi	Mickey Mouse, Tom and Jerry	Ms
Robert Phil	Goba	Dar to Lagos, Uncle JJ	Mr
Robert Phil	Goba	Pyaar Impossible	Mr

- The column **MOVIES RENTED** has multiple values.

DATABASE NORMALIZATION…

First Normal Form (1NF)

- 1NF rules
  - Each table cell should contain a single value.
  - Each record needs to be unique.

- The following are the steps of creating a table to be in the 1NF.
  - Identify the fields containing multiple information.
  - Re-design the table grouping similar information together.
  - Identify each set of related data with a primary key.

DATABASE NORMALIZATION…

First Normal Form (1NF)

ID	FULL NAMES	PHYSICAL ADDRESS	MOVIES RENTED	SALUTATION
1	Janet Jones	Mbezi	Mickey Mouse	Ms
2	Janet Jones	Mbezi	Tom and Jerry	Ms
3	Robert Phil	Goba	Dar to Lagos	Mr
4	Robert Phil	Goba	Uncle JJ	Mr
5	Robert Phil	Kinyerezi	Pyaar Impossible	Mr

DATABASE NORMALIZATION…

2NF

- A table is in a Second Normal form when it is in :-
  - First Normal Form.
  - It include no partial dependencies
    - ✓ no attribute is dependent on a portion of the composite primary key.

DATABASE NORMALIZATION…

2NF

- The followina table should normalized to 2NF.

*EmployeeID	LastName	FirstName	*ProjectNumber	ProjectTitle
EN1-26	O'Brien	Sean	30-452-T3	STAR manual
EN1-26	O'Brien	Sean	30-457-T3	ISO procedures
EN1-26	O'Brien	Sean	31-124-T3	Employee handbook
EN1-33	Guya	Amy	30-452-T3	STAR manual
EN1-33	Guya	Amy	30-482-TC	Web Site
EN1-33	Guya	Amy	31-241-TC	New catalog
EN1-35	Baranco	Steven	30-452-T3	STAR manual
EN1-35	Baranco	Steven	31-238-TC	STAR prototype
EN1-36	Roslyn	Elizabeth	35-152-TC	STAR pricing

DATABASE NORMALIZATION…

2NF

- The asterisks indicate the fields that make up the primary key of the table.
- A multi-field primary key is necessary because neither the EmployeeID nor the ProjectNum fields contain unique values.
- The reason there are repeated values in LastName, FirstName, and ProjectTitle is that these fields are dependent on only part of the primary key.
- In order to remove the partial dependency, you have to break the table into smaller tables.

DATABASE NORMALIZATION…

2NF

- Procedures of creating second normal form
  - To decide what fields belong together in a table, think about which field determines the values in other fields.
    - Create a table for those fields and enter the sample data.
  - Think about what the primary key for each table would be and about the relationship between the tables.
    - Make sure that you don't have repeating data in non-key fields.

DATABASE NORMALIZATION…

2NF

- The solution of the above table is:-

Employees

*EmployeeID	Last Name	First Name
EN1-26	O'Brien	Sean
EN1-33	Guya	Amy
EN1-35	Baranco	Steven
EN1-36	Roslyn	Elizabeth
EN1-38	Schaaf	Carol
EN1-40	Wing	Alexandra

Projects

*ProjectNum	ProjectTitle
30-452-T3	STAR manual
30-457-T3	ISO procedures
30-482-TC	Web site
31-124-T3	Employee handbook
31-238-TC	STAR prototype
31-238-TC	New catalog
35-152-TC	STAR pricing
36-272-TC	Order system

DATABASE NORMALIZATION…

2NF

- The solution of the above table is:-
  - Employees_Projects

*EmployeeID	*ProjectNum
EN1-26	30-452-T3
EN1-26	30-457-T3
EN1-26	31-124-T3
EN1-33	30-328-TC
EN1-33	30-452-T3
EN1-33	32-244-T3

DATABASE NORMALIZATION…

3NF

- Conditions for third normal form:-
  - A table must be second normal form (2NF).
  - There are no transitive dependencies.

- A transitive dependency is a type of dependency in which the value in a non-key field is determined by the value in another non-key field and that field is not a candidate key.

DATABASE NORMALIZATION…

3NF

- Consider a table below with a single field primary key and repeating values in non-key fields.

*ProjectNum	ProjectTitle	ProjectMgr	Phone
30-452-T3	STAR manual	Garrison	2756
30-457-T3	ISO procedures	Jacanda	2954
30-482-TC	Web site	Friedman	2846
31-124-T3	Employee handbook	Jones	3102
31-238-TC	STAR prototype	Garrison	2756
31-241-TC	New catalog	Jones	3102
35-152-TC	STAR pricing	Vance	3022
36-272-TC	Order system	Jacanda	2954

- The phone number is dependent on the project manager (a transitive dependency).

DATABASE NORMALIZATION…

3NF

- Procedures of creating third normal form
  - Think about which fields belong together and create new tables to hold them.
  - Enter the sample data and check for unnecessarily (not part of primary key) repeated values.
  - Identify the primary key for each table and, if necessary, add foreign keys.

DATABASE NORMALIZATION…

3NF

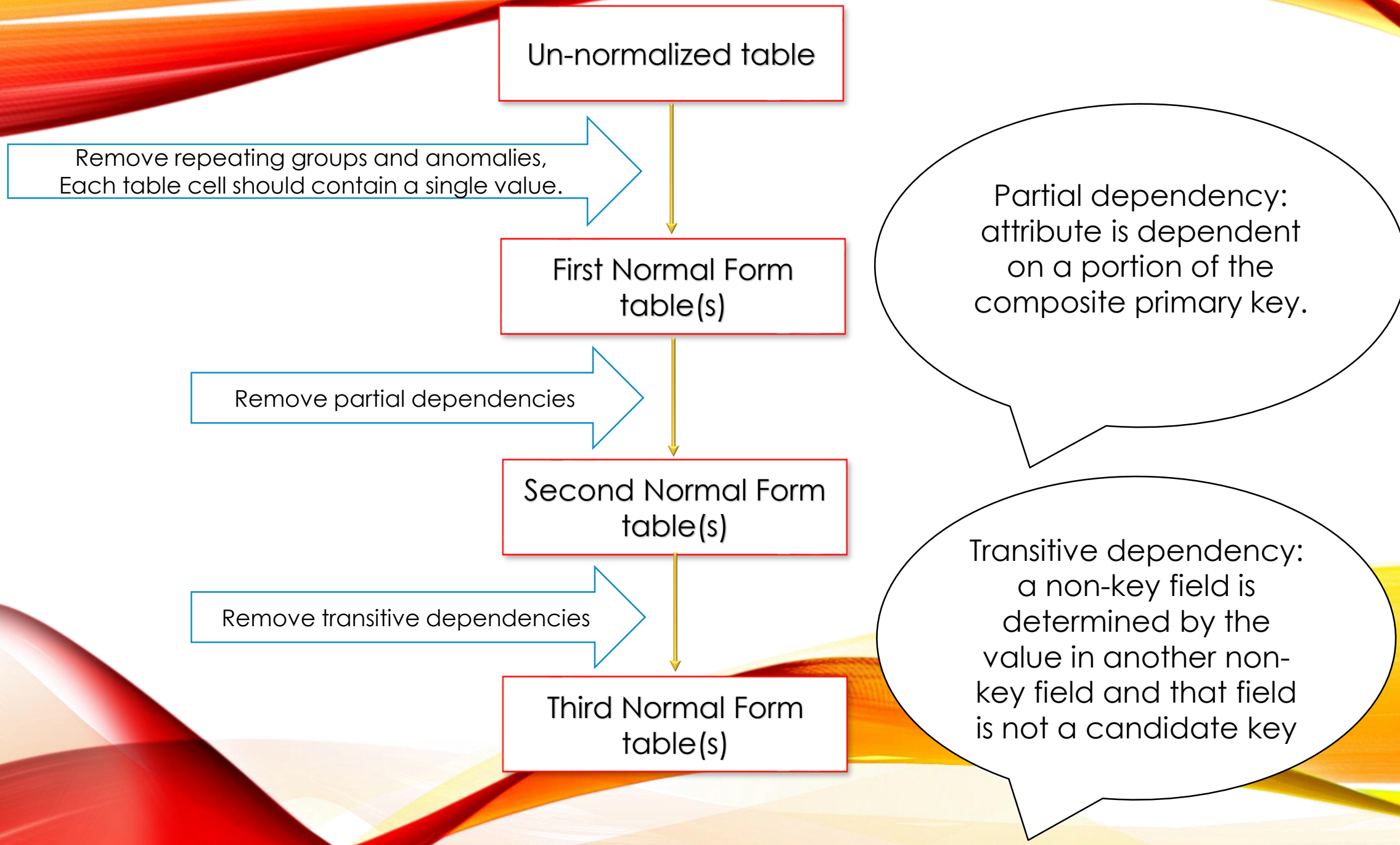
Projects

*ProjectNum	ProjectTitle	ProjectMgr
30-452-T3	STAR manual	Garrison
30-457-T3	ISO procedures	Jacanda
30-482-TC	Web site	Friedman
31-124-T3	Employee handbook	Jones
31-238-TC	STAR prototype	Garrison
31-241-TC	New catalog	Jones
35-152-TC	STAR pricing	Vance
36-272-TC	Order system	Jacanda

Managers

*ProjectMgr	Phone
Friedman	2846
Garrison	2756
Jacanda	2954
Jones	3102
Vance	3022

In most cases 3NF should be sufficient to ensure that your database is properly normalised, however higher normal forms can be achieved.


Un-normalized table
Remove repeating groups and anomalies,
Each table cell should contain a single value.
First Normal Form table(s)
Remove partial dependencies
Second Normal Form table(s)
Remove transitive dependencies
Third Normal Form table(s)
Partial dependency: attribute is dependent on a portion of the composite primary key.
Transitive dependency: a non-key field is determined by the value in another non-key field and that field is not a candidate key

"Don't judge each day by the harvest you reap but by the seeds that you plant." – Robert Louis Stevenson