**CHAPTER NINE**

**DATA NORMALIZATION**

**Introduction**

Normalization is the process of applying a number of rules to the tables, which have been identified in order to simplify.

**Terminology**

**Normalization:** Process of eliminating redundancies through creation of more tables.

**Normal forms:** Rules used in normalization

**Normal Forms**

The first three rules applied to normalization are referred to as: -

• First Normal Form (1NF)

• Second Normal Form (2NF)

• Third Normal Form (3NF)

**First normal form**

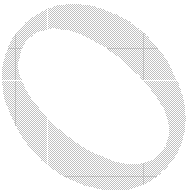
A table or relation is said to be in first normal form, if and only if it contains non-repeating groups i.e. it has no repeated values for particular attributes in a simple record (i.e. no more than one value may be contained in each field). If there are repeating groups and attributes they should be isolated to form a new entity.

**Second normal form**

A table is said to be in 2NF if and only if it is in 1NF and every non- key attribute is fully dependent on the key attributes they should be isolated to form a new entity.

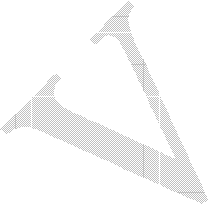
**Third normal form**

A table is said to be in 3NF if and only if it is 2nd NF and every non-key attribute is not dependent on any other non-key attribute. All non-key attributes that are dependent on other non-key attributes, should be isolated to form a new entity











**InvoiceNo**.\_

Date

Customer Delivery to Address

Product Code Description Quantity Price Amount

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

Thank you.

Amount\_

*: A customer’s Invoice*

Un-normalized data

Invoice (Invoice no., Date, Customer, Cust\_address, Deliv\_To, Product code, Quantity, Chapter Price, amount, Invoice amount)

INF (Identify and separate repeating groups to form a new entity) INVOICE (Invoice number, date customer address, Deliv\_

address, Invoice\_Amount)

PRODUCT (Product code, invoice number, product description, Chapter price, amount)

2NF (identity and separate non-key attributes not fully dependent on key attribute)

INVOICE (Invoice-no, date, customer address, del.address, invoice total)

INVOICE PRODUCT (prod-code, prod-description, Chapter price)

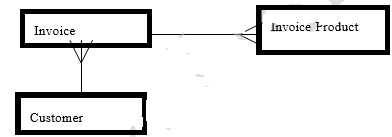
3NF (Identify non-key attributes dependent on other non-key attributes)

INVOICE (Invoice-no, Customer\_Number, Prod-code, date, invoice total)

INVOICE PRODUCT (Prod-code, prod-description, Chapter price) CUSTOMER (Customer\_Number, Customer\_Name,

customer\_Address, del.address)

NB: Whenever there is no composite key the table is in 3NF



An ERD for the Customer’s Invoice

**Advantages and Disadvantages of Normalization**

**Advantages**

1. It is a formal technique with each stage of normalization process elimination a particular type of undesirable dependency as well as each stage of normalization eliminating a certain type of redundancy.

2. It highlights constraints and dependencies in the data and helps in understanding the nature of the data.

3. The 3NF produces well-designed databases, which provide a high degree of independence.

**Disadvantages**

1. It depends a thorough understanding of the entities and their relationships.

2. It’s a complex process particularly if the entities are many.