Introduction to RDBMS Concepts

Introduction to Initial Data Storage Methods

- Data was represented in one or more flat files
- Flat file is nothing but electronic document structure for data storage.
- Every business group has its own set of files

Disadvantages of Flat File System

- No centralized control.
- Data Redundancy
- Data Inconsistency
- Data can not be shared
- Standards can not be enforced
- Security issues

Data Models

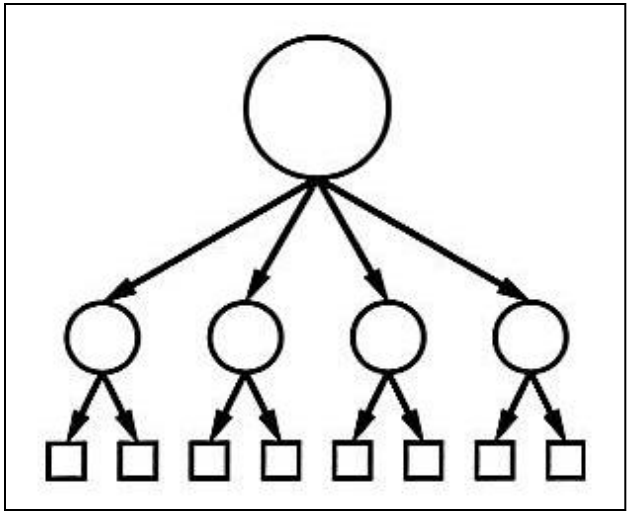
- Data Model is a collection of concepts that can be used to describe the structure of a database.

- Data Models are categorized into:

  - Hierarchical Model
  - Network Model
  - Relational Model

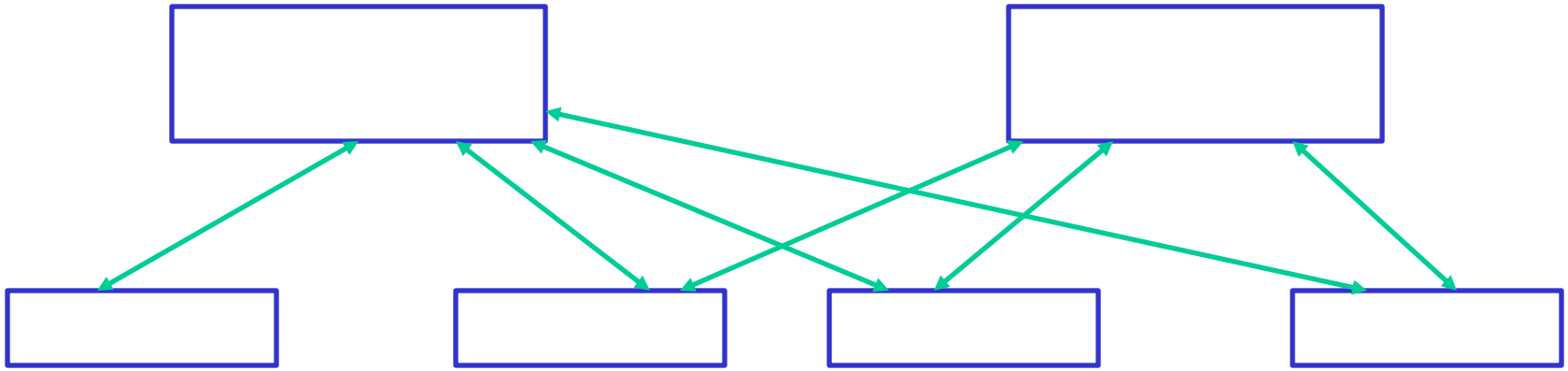
Hierarchical Model

- Organizes data into tree structure.
- Hierarchy of parent-child data segment.
- Child data segments can have repeating information.
- Data is in series of records which has set of field values attached to it.



Network Model

- Data were modeled with more than one parent per child.
- Network model permitted many to many relationships in data.



Relational Model

- Relational database allows the definition of:
  - Data structures
  - Storage and retrieval operations and
  - Integrity constraints

- Data and relations between them are organized in tables.
- This model is developed by E. F. Codd.

Empno	Empname	Desig	Salary	Deptno
E001	Vandana	Manager	20000	10
E002	Amit	Executive	14000	20
E003	Amit	Accountant	10000	30
E004	Anju	Clerk	9000	10

What is Database?

- Database is an organized collection of related information.
- It has some inherent meaning
- Represents some real world aspects.
- Designed, built and populated with data for a specific purpose.

Database Management System

- The system which allows the user to:
  - Add new data
  - Modify existing data
  - Remove unnecessary data
  - Read the data

- Known as DBMS

Advantages of DBMS

- Centralized control.
- No Data Redundancy
- Data Consistency
- Data can be shared
- Security can be enforced
- Integrity can be maintained
- Data independence
- Provides multiple user interfaces
- Provides backup and recovery.

Database Schema

- Structure described in a formal language supported by the DBMS.

- Refers to the organization of data to create a blueprint of how a database will be divided into database tables.

Data Instances

- The database instance refers to a complete database environment.

- It includes RDBMS software, table structure, stored procedures and other functions.

Data Independence

- Data independence means that the application is independent of the :
  - storage structure and
  - access strategy of data.

- In other words, The ability to modify the schema definition in one level should not affect the schema definition in the next higher level.

- Data independence is thought in two levels:
  - Logical level
  - Physical level

Database Languages

- SQL is used for querying the database.
- It has been classified into:
  - DDL – Data definition language which is used to describe data and data structure.
  - DML – Data manipulation language which is used for the operations such as search, read, change, store data etc.
  - DQL –Data Query Language which contains only one command – SELECT , used to retrieve the data.
  - DCL – Data control language which is used to access or deny user privileges.
  - TCL – Transact control language which allows the statements to be grouped together into logical transactions.

What is RDBMS?

- Abbreviation for Relational Database Management System
- RDBMS Data is structured in:
  - Database tables
  - Fields
  - Records
- RDBMS table consists of table rows.
- Each table row consists of one or more fields.

RDBMS continued…

- Relational table has the following properties:
  - Values Are Atomic
  - Each Row is Unique
  - Column Values Are of the Same Kind
  - The Sequence of Columns is Insignificant
  - The Sequence of Rows is Insignificant
  - Each Column Has a Unique Name

Relational Database continued…

- Relational database model is a collection of tables.

Table

DEPT	
Deptno	Dname

Attribute/
Field/
Column

Table

EMP				
Empno	Ename	Desig	Salary	Deptno

Relational database

- A relational table is a flat file which is composed of :
  - A set of named columns and
  - Arbitrary number of rows.

In a relational database the data is maintained across multiple tables.

Record/ Tuple/ Row

EMP				
Empno	**Ename**	**Desig**	**Salary**	**Deptno**
E001	Vandana	Manager	20000	10
E002	Amit	Executive	14000	20
E003	Amit	Accountant	10000	30
Attributes				

Relationship

DEPT

Deptno	Dname
10	Admin
20	IT
30	Account

Attributes

A relationship is an association between two or more tables.

EMP

Empno	Ename	Desig	Salary	Deptno
E001	Vandana	Manager	20000	10
E002	Amit	Executive	14000	20
E003	Amit	Accountant	10000	30

Attributes

Keys

- Keys are fundamental to the concept of relational database
- They enable tables in the database to be related with each other.
- Types of keys are:
  - Primary key
  - Candidate key
  - Alternate key
  - Foreign key

Primary key

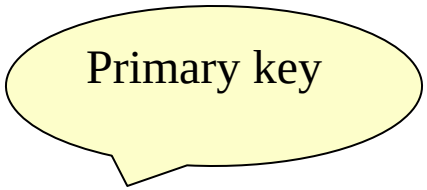
- Consider the following table:

Empname	Desig	Salary	Deptno
Vandana	Manager	20000	10
Amit	Executive	14000	20
Amit	Executive	14000	20
Anju	Clerk	9000	10

- In this table, there exists two employees with the same data.
- It means the same data is repeated here.
- Uniqueness of the data is not maintained.
- To avoid this and maintain uniqueness of the data **Primary key** is used.

Primary key continued…

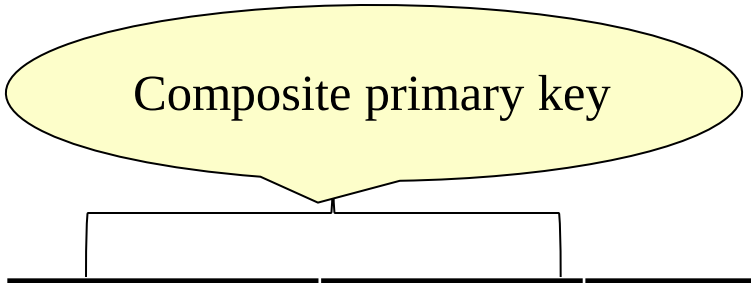
- In this table the employee is identified by it's Empno.
- Empno is used here to uniquely identify the Empno.
- Hence Empno is considered here as a primary key.



Empno	Empname	Desig	Salary	Deptno
E001	Vandana	Manager	20000	10
E002	Amit	Executive	14000	20
E003	Amit	Accountant	10000	30
E004	Anju	Clerk	9000	10

Composite primary key

- Primary key consists of two or more than two columns is considered to be composite primary key.



Invoice No	ProductId	Qnty_shpd	Discount
153	66	3000	16.7
153	26	6000	16.7
153	91	6000	16.7
154	66	3000	16.7
154	26	6000	16.7
154	91	6000	16.7

Candidate key

- The key which is in the race of primary key is candidate key.
- Each table may have one or more than one as a candidate key.
- One of the candidate key is selected as primary key.
- For e.g:

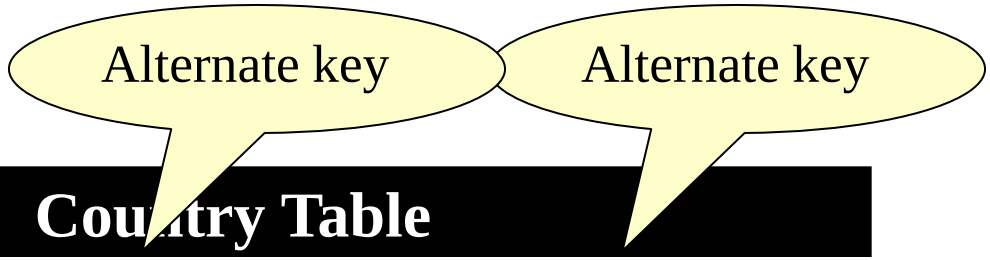
Candidate key

Country Table

Country_Code	Country_Name	Capital
IN	India	Delhi
AU	Australia	Canabara
US	Unites States	Washington DC

Alternate key

- An alternate key is similar to a primary key.
- It accepts null values; where as the primary key does not.
- The null values can be submitted to the attribute in a tuple.

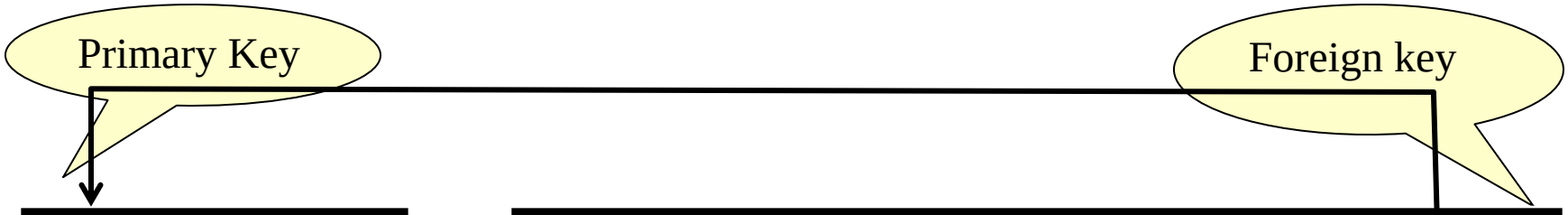


Country Table

Country_Code	Country_Name	Capital
IN	India	Delhi
AU	Australia	Canabara
US	Unites States	Washington DC

Foreign key

- Foreign key is a link or relationship between two tables.
- Ensures that the data stored in a database is consistent.



DEPT

Deptno	Dname
10	Admin
20	IT
30	Account
Attributes	

EMP

Empno	Ename	Desig	Salary	Deptno
E001	Vandana	Manager	20000	10
E002	Amit	Executive	14000	20
E003	Amit	Accountant	10000	30
Attributes				

Database Design

- Requirement analysis.
- Diagrammatic representation.
- Translating diagrams to tables.
- Refining tables based on fixed set of rules.

Requirement analysis

- Requirement analysis is the process of determining user expectations.
- Requirements must be quantifiable, relevant and detailed.
- Requirement analysis involves frequent communication with users.
- Once the requirements are gathered from the users, next step is to represent it in the diagrammatical format.

Database Modeling

- Conceptual representation of the data structures required by a database.

- The goal of the data model is to make sure that the all data objects required by the database are completely and accurately represented.

Data Modelling Methodology

- Methodology used to create a data model :
  - Entity-Relationship (ER) approach.

Entity Relationship Modelling

- Entity-Relationship (E-R) Modeling is a conceptual modeling tool.

- Perceives the business environment in terms of participating “entities” and the “relationship” between them.

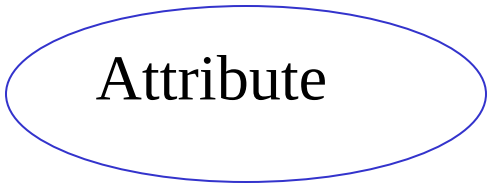
Entity (set)

- It is a “data object”.
- Entity is a thing in the real world with an independent existence.
- Represented by a rectangle shape containing Entity name.

Attributes

- Characteristics of an entity that provides descriptive details of it.

- Every attribute must be given a unique name.



Attributes continued…

- Attributes are classified as followed:

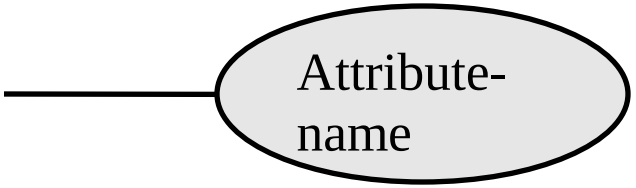
  - **Composite attribute** is an attribute that can be sub divided further into additional attributes. For ex. The Attribute Empname can be subdivided into First_Name and Last_Name

  - **Single-valued attribute** can have only a single value. For ex. A Employee can have only one EmpId. A person's age

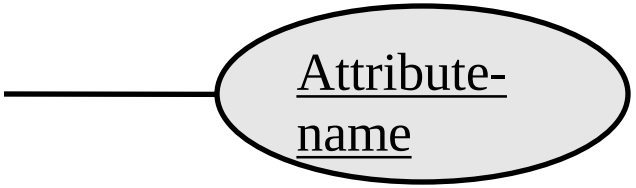
  - **Multi valued attributes** can have many values. For instance a Employee may have several college degrees as his qualification, phone numbers, Email-d etc..

The Chen notation attributes

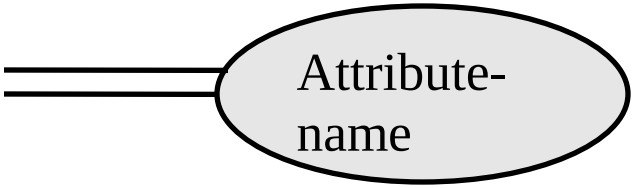
- The Chen notation uses following symbols:
- Simple Attribute



- Primary key

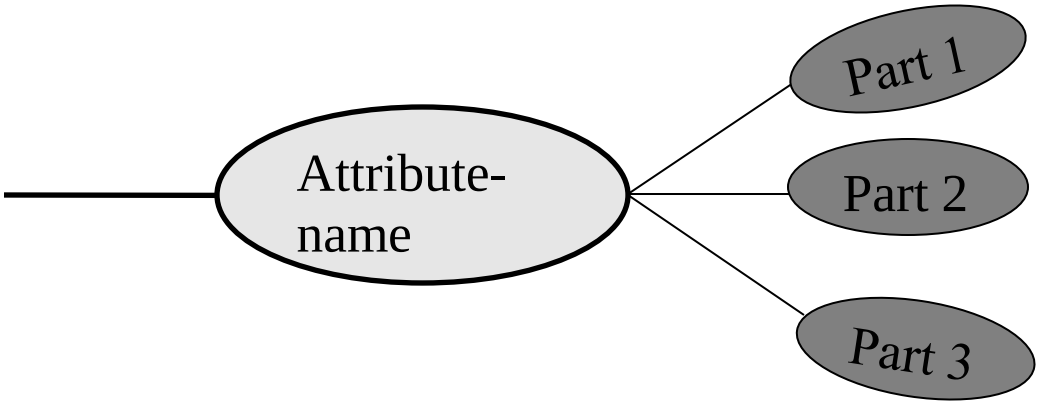


- Multi-valued attribute

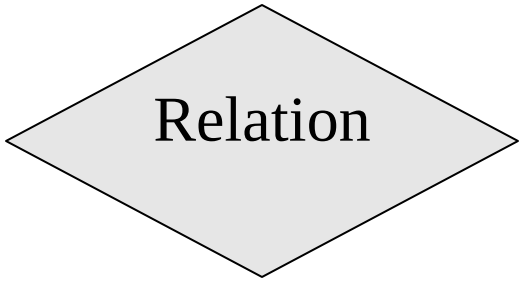


The Chen notation attributes continued…

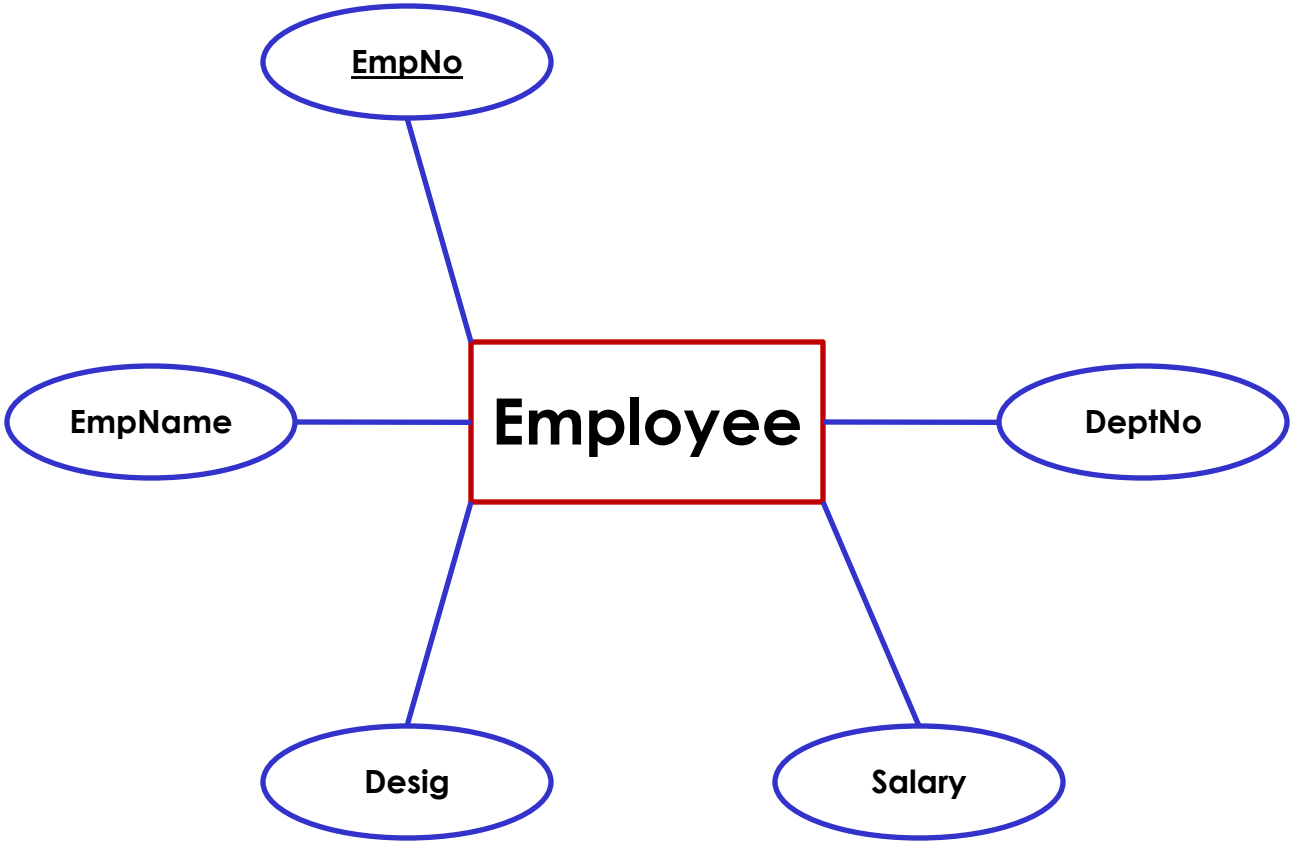
- Composite Attribute



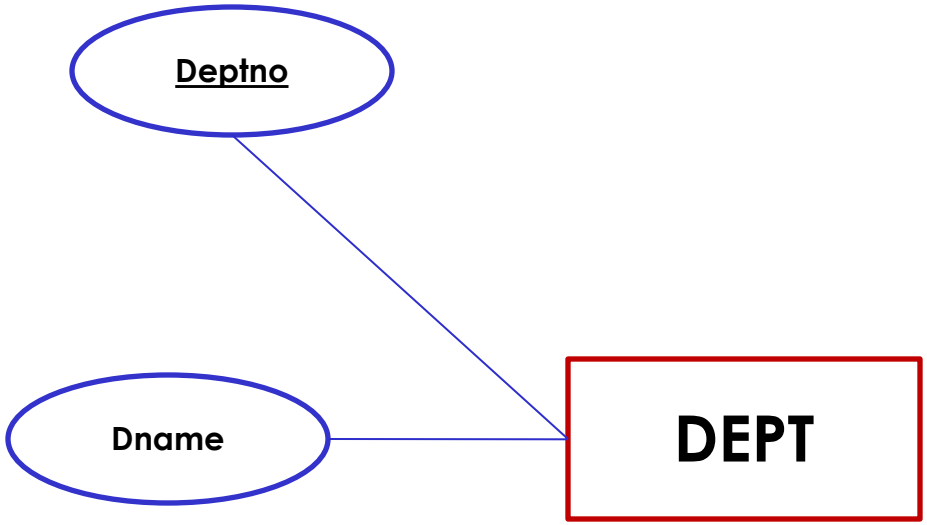
- Relationship



E-R diagram for EMP table



E-R diagram for DEPT table

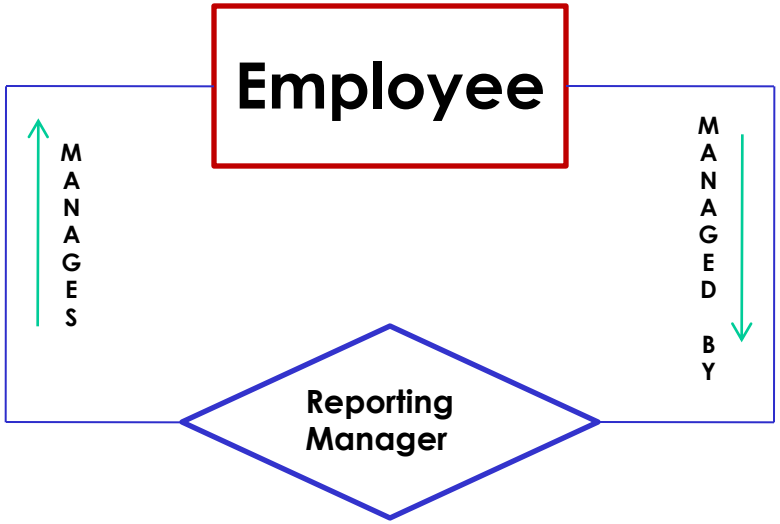


Relationship's Degree

- Indicates the number of entities involved in the relationship.
- It can be classified as follows:
  - Unary relationship
  - Binary relationship
  - Ternary relationship

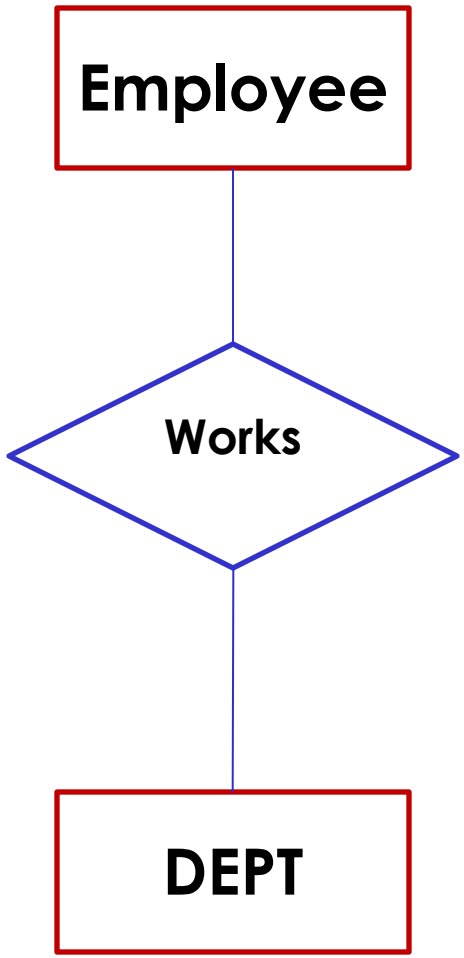
Relationship's Degree - Unary

Unary relationship exists when an association is maintained within a single Entity.



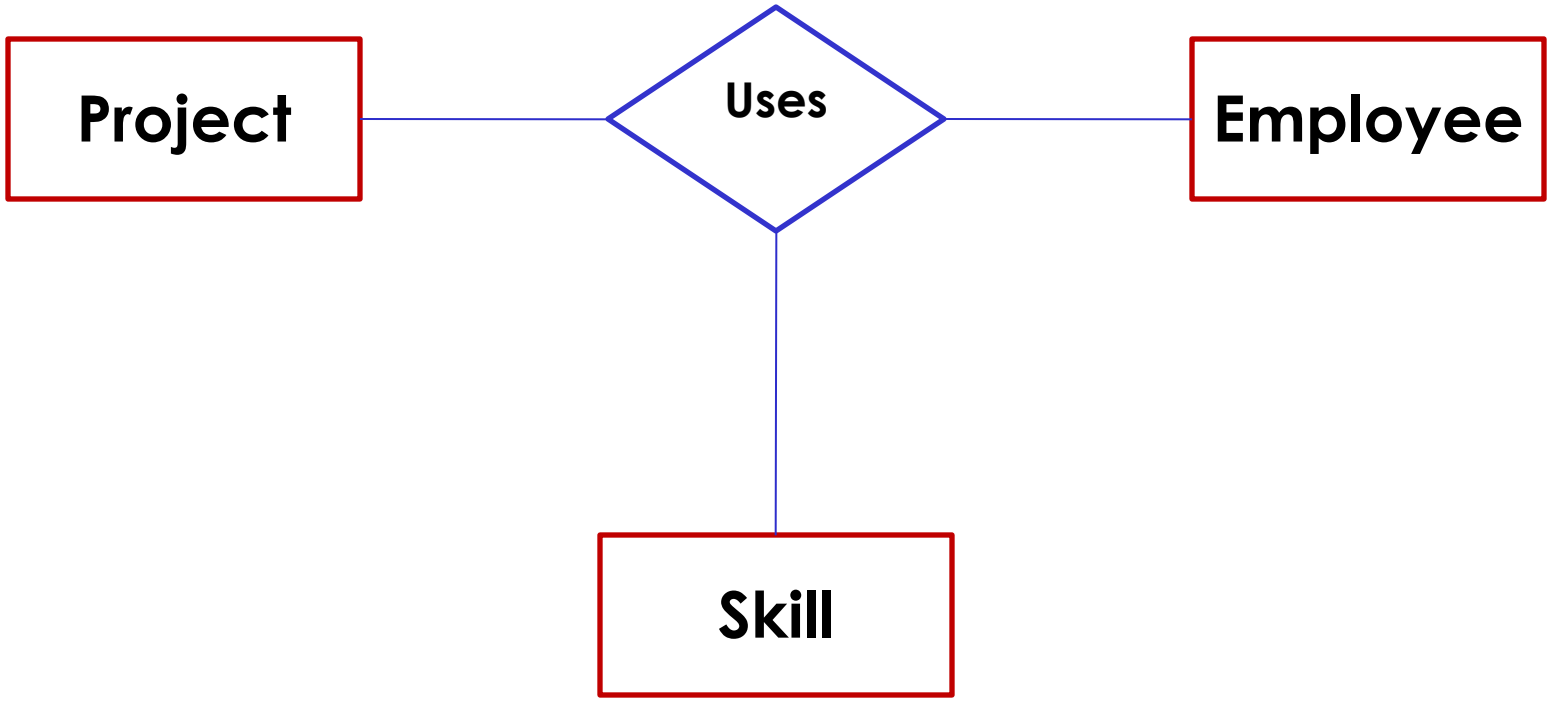
Relationship's Degree - Binary

Binary relationship exists when two entities are associated.



Relationship's Degree - Ternary

Ternary relationship exists when three entities are associated.



Cardinality

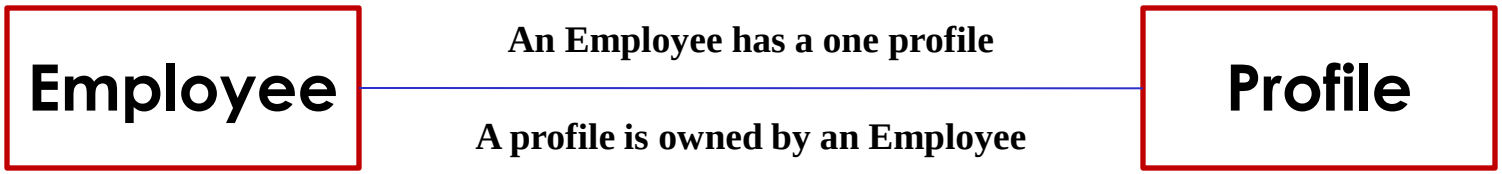
- Cardinality means specific number, or range of numbers, of entities involved in a relationship.

- Cardinality is usually described as:

  - One-to-one
  - One-to-many
  - Many-to-many

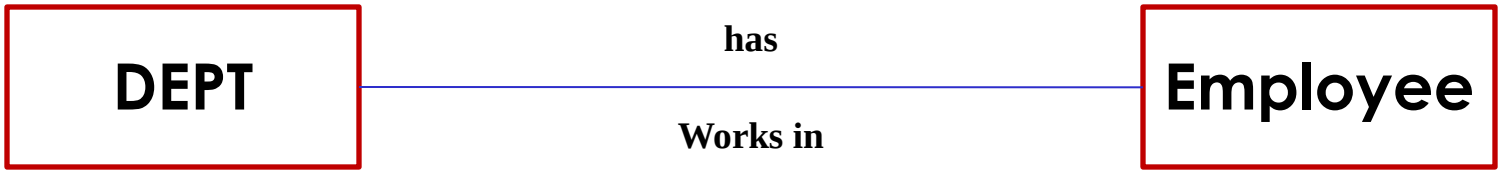
Cardinality - One to one

- One occurrence of an entity relates to only one occurrence in another entity.
- Bidirectional relationship which means single valued at both the side.



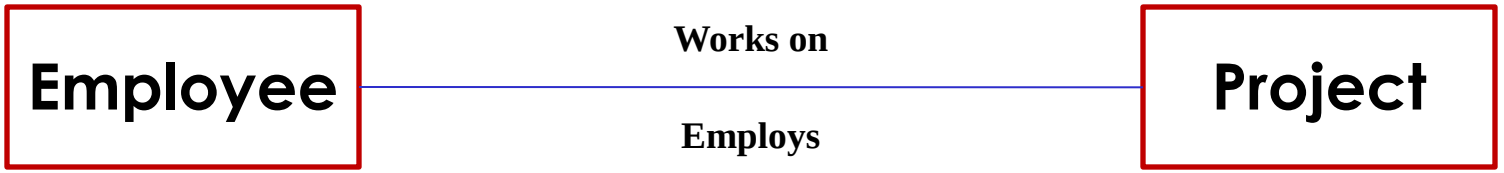
Cardinality – One to Many

- One occurrence in an entity relates to many occurrences in another entity.

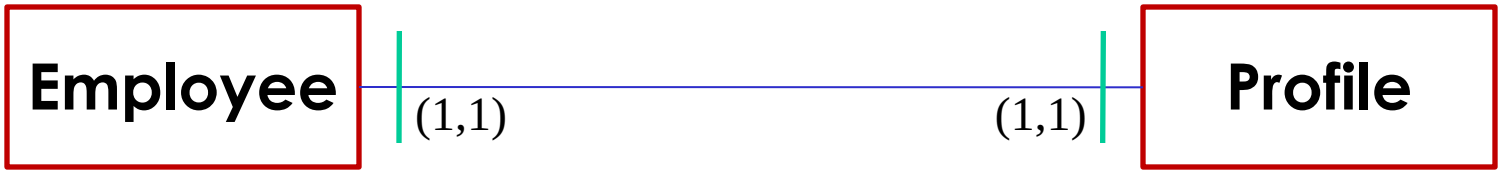


Cardinality – Many to Many

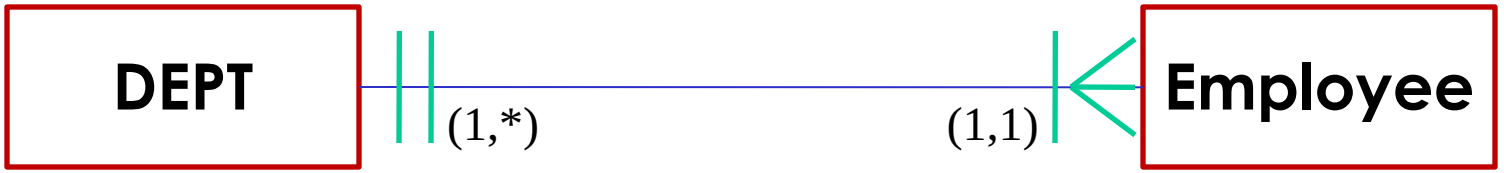
- Many occurrences in an entity relate to many occurrences in another entity.



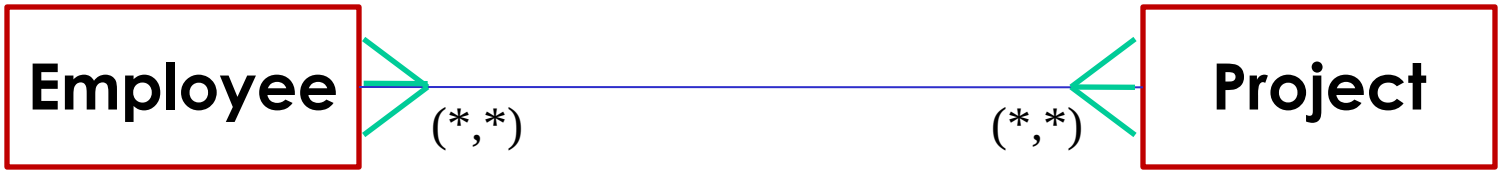
One to One cardinality



One to Many - Cardinality



Many to Many - Cardinality

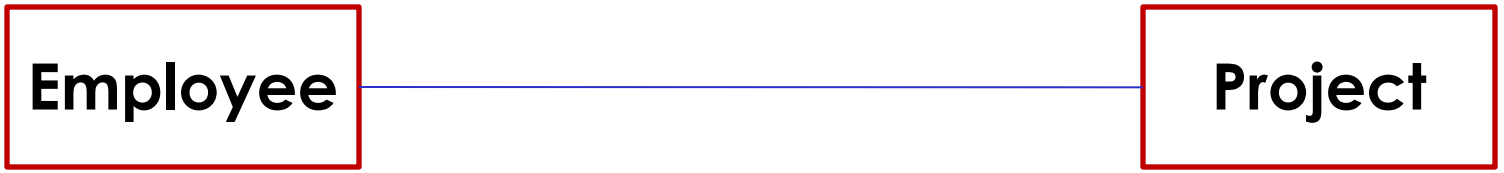


Entity types

- Strong entity
- Weak entity
- Recursive entity
- Composite entity

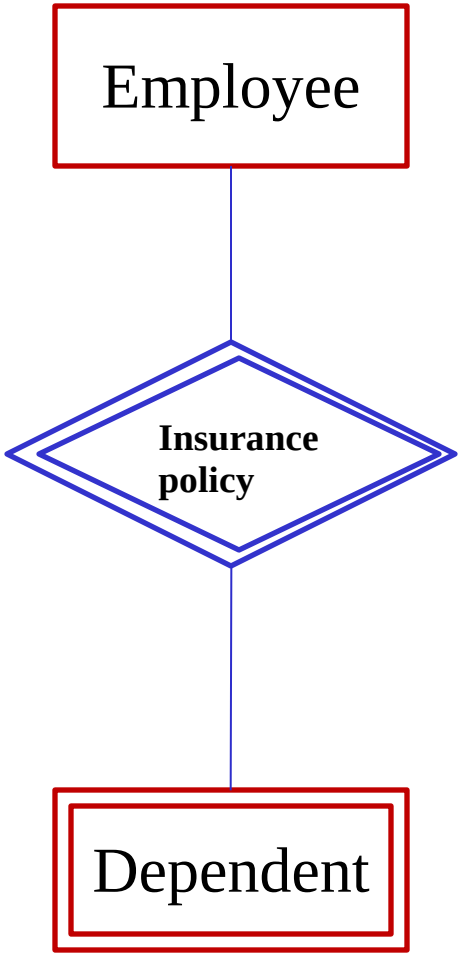
Strong entity

- An entity set that has a primary key is termed as **strong entity set.**



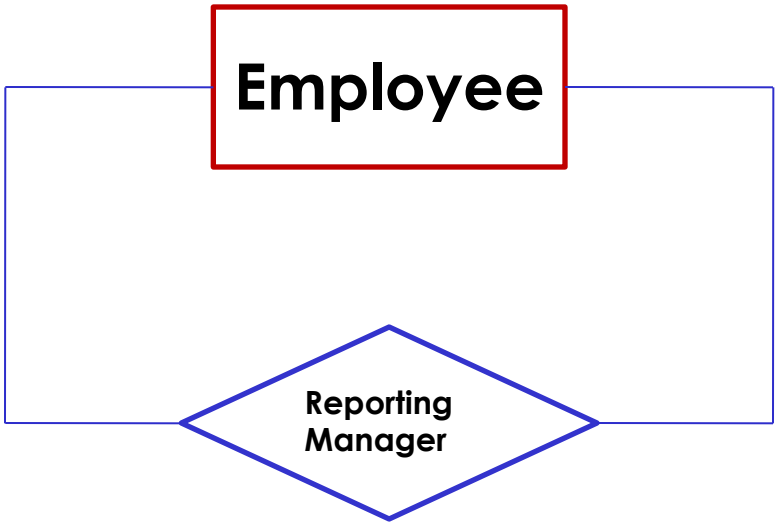
Weak entity

- An entity set that does not have sufficient attributes to form a primary key is termed as a weak entity set.



Recursive entity

- Relation exists between occurrences of the same entity set.
- This occurs in a unary relationship.



Composite entity

- Bridge entity composed of the primary keys of each entity.
- Bridge entity is known as composite entity.
- Represented by a diamond shape with in a rectangle in ER diagram.

Constraints

- Constraints are the rules enforced on the data stored in the database.
- Constraints can be applied to already existing tables as well as while at creation time.
- Constraints can be applied at the column level as well as the table level.

Type of Constraints

- Not null
- Unique
- Primary key
- Foreign key
- Check

The NOT NULL Constraint

- If one wants to ensure that a record is not added to a table unless some specific fields have some value, then NOT NULL constraint is used.

Empname column must have value so this column will be NOT NULL

Empno	Empname	Desig	Salary	Deptno
E001	Vandana	Manager	20000	10
E002	Amit	Executive	14000	20
E003	Amit	Accountant	10000	30
E004	Anju	Clerk	9000	10

The UNIQUE Constraint

- UNIQUE constraint can be used to ensure that no two records in the column have the same value.

- But you can have one null value for a table.

The Primary key Constraints

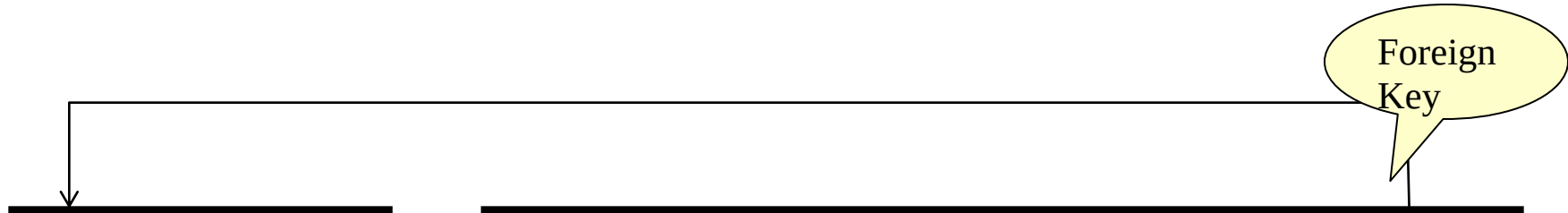
- No two records can have the same value for the column.
- NOT NULL is automatically applied to the column. That is no record can contain a null value for the column.
- An index is created on the column. This results into quicker retrieval of records.

Primary Key

Empno	Empname	Desig	Salary	Deptno
E001	Vandana	Manager	20000	10
E002	Amit	Executive	14000	20
E003	Amit	Accountant	10000	30
E004	Anju	Clerk	9000	10

The REFRENCES Constraints

- Values of the foreign key
  - must be either null, or
  - if non-null, must match with the primary key value of some record of the Master relation.

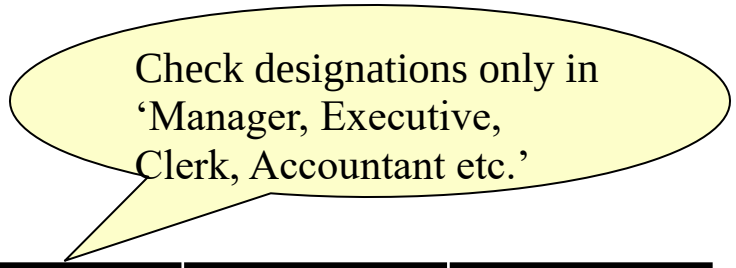


DEPT	
Deptno	**Dname**
10	Admin
20	IT
30	Account
Attributes	

EMP				
Empno	**Ename**	**Desig**	**Salary**	**Deptno**
E001	Vandana	Manager	20000	10
E002	Amit	Executive	14000	20
E003	Amit	Accountant	10000	30
Attributes				

Check constraint

- Restrict a column value to a set of values defined by the constraint.



Empno	Empname	Desig	Salary	Deptno
E001	Vandana	Manager	20000	10
E002	Amit	Executive	14000	20
E003	Amit	Accountant	10000	30
E004	Anju	Clerk	9000	10

Normalization

- Normalization is a design technique that is widely used as a guide in designing relational databases.

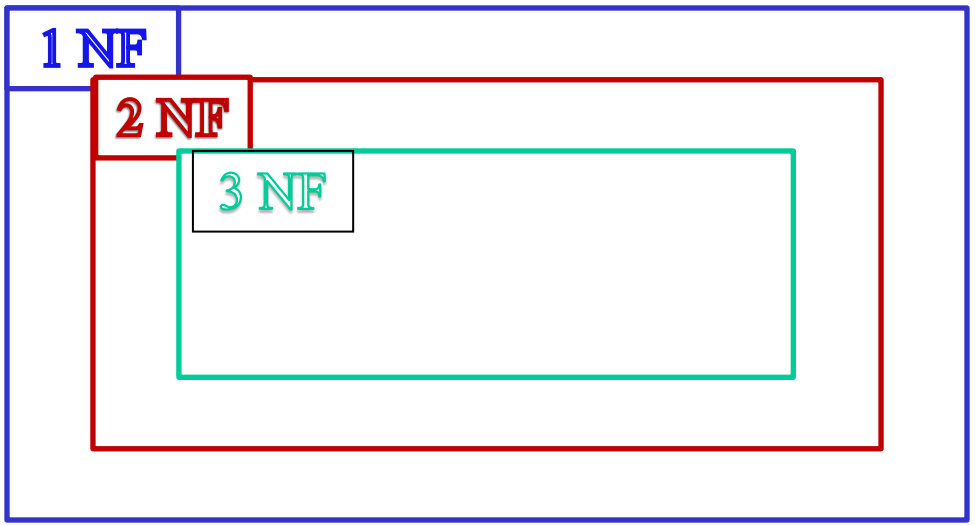
- The goal of normalization is to create a set of relational tables that are free of redundant data and that can be consistently and correctly modified.

- Normalization theory is based on the concepts of normal forms.

Normal Forms

- A relational table is said to be a particular normal form if it satisfies a certain set of constraints.

Hierarchy of Normal Forms



Normal Forms are **INCREMENTAL**

Unnormalized

Order_Details

Order #	Order Date	Item#	Qty	Price	Cust code#	Name	Address	City
245	2-Jan-2009, 15-Apr-2010	I768,I656	1,3	350,125	C1	Steel Co	Works bldg	Mumbai
246	31-Aug-2010, 5-Mar-2011	I780, I237	4, 8	320, 120	C2	Acme Corp	North wing	Pune

multi-valued columns

1NF

- Eliminate variable repeating fields and groups so that all attributes take atomic values.

A relation is said to be in "first normal form" (1NF) if and only if all its attributes assume only atomic values.

Unnormalized → 1 NF

In 1NF

Order #	Item#	Qty	Price	Cust code#	Name	Address	City
245	I768	1	350	C1	Steel Co	Works bldg	Mumbai
245	I656	3	125	C1	Steel Co	Works bldg	Mumbai
246	I780	4	320	C2	Acme Corp	North wing	Pune
246	I237	8	120	C2	Acme Corp	North wing	Pune

2NF

- Eliminate fields that are facts about only a ***subset*** of the key so that all non-key domains are fully functionally dependent on the primary key.

A relation is said to be in 2NF if and only if it is in 1NF and every non-key attribute is fully functionally dependent on the primary key.

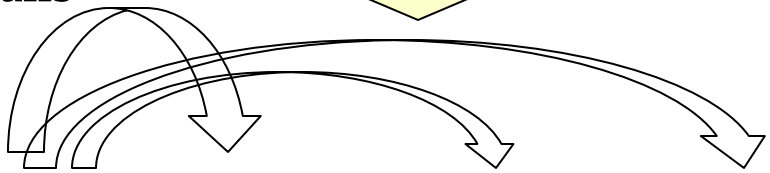
Why 2NF?

Order

Order#	Order Date
245	
246	

Order Details

Not fully functional dependency on primary key



Order #	Item#	Qty	Price	Cust code#	Name	Address	City
245	I768	1	350	C1	Steel Co	Works bldg	Mumbai
245	I656	3	125	C1	Steel Co	Works bldg	Mumbai
246	I780	4	320	C2	Acme Corp	North wing	Pune
246	I237	8	120	C2	Acme Corp	North wing	Pune

1NF ⟹ 2NF

Order

Order#	Order Date
245	
246	

Order Details

Order #	Item#	Qty	Price
245	I768	1	350
245	I656	3	125
246	I780	4	320
246	I237	8	120

Cust_Order Details

Order #	Cust code#	Name	Address	City
245	C1	Steel Co	Works bldg	Mumbai
246	C2	Acme Corp	North wing	Pune

Why 3NF?

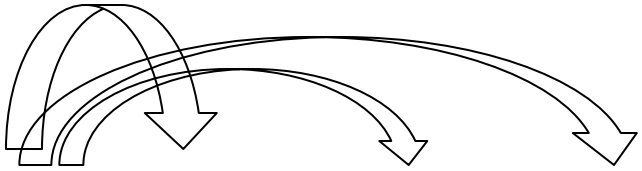
Order

Order#	Order Date
245	
246	

Order Details

Order #	Item#	Qty	Price
245	I768	1	350
245	I656	3	125
246	I780	4	320
246	I237	8	120

Cust_Order Details



Order #	Cust code#	Name	Address	City
245	C1	Steel Co	Works bldg	Mumbai
246	C2	Acme Corp	North wing	Pune

Transitive dependency. Custname, Address and city are transitively dependant .

2NF ⟹ 3NF

- Eliminate non-key fields that are transitively dependant on key fields.

3NF

Order

Order#	Order Date
245	
246	

Order Details

Order #	Item#	Qty	Price
245	I768	1	350
245	I656	3	125
246	I780	4	320
246	I237	8	120

Cust_Order Details

Order #	Cust code#
245	C1
246	C2

Cust Details

Cust code#	Name	Address	City
C1	Steel Co	Works bldg	Mumbai
C2	Acme Corp	North wing	Pune

Security

- Keeps control on the data accessed by different users.
- The different users are decided.
- Privileges are set on different roles.
- Roles are groups of users.
- The privileges are given to different users.

Data Control Language

- Adds security features within the database
- Controls the data accessed by the different users.
- DBA can have complete control on each and every operation performed on the database by all the users of the Database.
- Complete data security can be implemented.

Data Control Language continued…

- It has two commands:
  - Grant Command
  - Revoke Command

Transaction Processing

- A transaction is a sequence of one or more SQL statements that together form a logical unit of work.
- Each statement in the transaction performs some part of a task.
- Every statement is required to complete the task.

Need for Transaction Management

- If transaction management system is not available then

  - In case of problems like machine failure it is difficult to check whether updation has taken place for a table or not.
  - Concurrency control and recovery.
  - Serializability : When several concurrent transactions are trying to access the same data item, the instructions within these concurrent transactions must be ordered in some way so as there are no problem in accessing and releasing the shared data item.
  - Prevents reading Inconsistent Data

Steps in Transaction Management

- All the changes to the data are made permanent within a Database.
- None of the changes are made permanent within a Database.
- If the transaction stops in the middle because of failure of operating system or user program then
  - Database is automatically restored to the state in which it was before starting the process.

Transaction Properties

- Atomicity
- Consistency
- Isolation
- Durability

Atomicity

- A transaction is an atomic unit of processing.
- It is either performed entirely or not performed at all.

Consistency

- Its said to be consistent.

- It preserves its complete execution which takes the database from one consistent state to another.

Isolation

- The execution of a transaction should not be interfered with by any other transactions executing concurrently.

Durability

- The changes applied to the database by a committed transaction must persist in the database.

- These changes must not be lost because of any failure.

Transaction Control Language

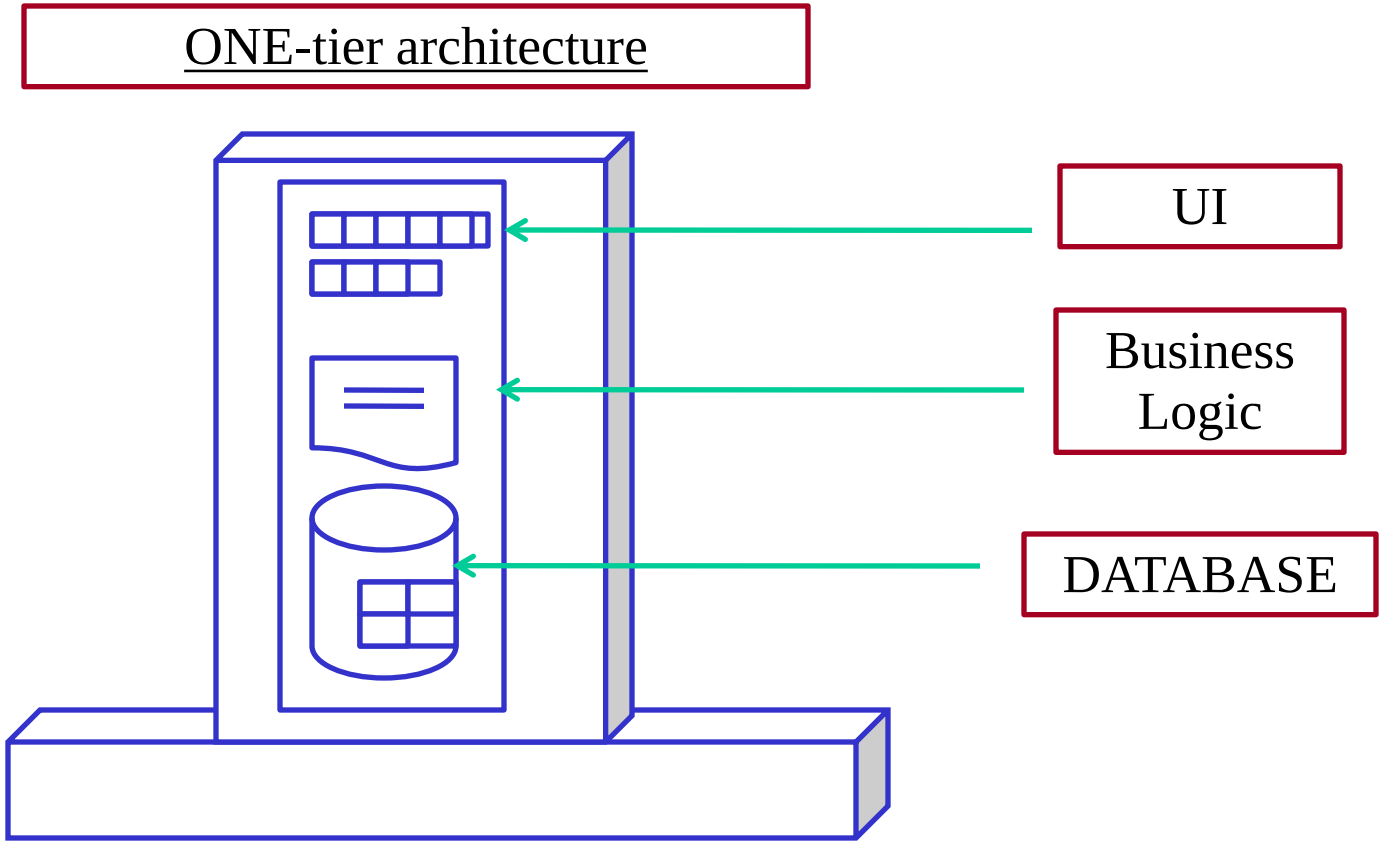
- Proper transaction management can be implemented using TCL.

- The TCL commands are:
  - Commit
  - Rollback

Distributed Databases and client-Server Computing

- The Client-Server software architecture is a versatile, message-based and modular infrastructure that is intended to improve:
  - Usability
  - Flexibility
  - Interoperability
  - Scalability

One-tier Architecture



Two-tier Computing