Chapter 1

* Database: organized collection of logically related data
* Data: stored representations of meaningful objects and events 🡪 structured and un---
* Information: data processed to increase knowledge in the person using the data
* Metadata: data that describes the properties and context of user dat
* Disadvantages of file processing 🡪 this is the advantages of data base
* Program-Data Dependence 🡪 each program
* maintain his/her own data
* needs to include code for the metadata
* Non-standard file formats
* Duplication of Data
* Waste of space
* Data changes in one file could cause inconsistencies
* Limited Data Sharing
* Lengthy Development Times
* Excessive Program Maintenance
* Database cost and risk
* New, specialized personnel
* Installation and management cost
* Conversion costs
* Need for explicit backup and recovery
* Database Management System: A software system that is used to create, maintain, and provide controlled access to user databases
* Elements of database: data models – relational databases – use of internet – database applications 🡪 programs do activities create – delete – update database
* Data Warehouse: Integrated decision support system derived from various operational databases
* Enterprise Resource Planning (ERP): Integrate all enterprise functions
* Range of data base applications: personal – workgroup – departmental - enterprice
* Components of data base environment:
* CASE Tools–computer-aided software engineering
* Repository–centralized storehouse of metadata
* Database Management System (DBMS) / Database
* Application Program / User Interface
* Data/Database Administrators–personnel responsible for maintaining the database
* System Developers–personnel responsible for designing databases and software
* End Users

Chapter 2

* Enterprise Data Model: first step – specify scope – specify entity types and relations
* Information Systems Architecture ISA: Conceptual blueprint for organization’s desired information systems structure
* Data - ERD
* Processes – DFD
* Data Network–topology diagram
* Events
* Reasons for events and rules
* CASE: Computer-Aided Software Engineering –software tools providing automated support for systems development
* Data modeling / Code generation / Repositories
* Packaged Data Models: Model components that can be customized, and assembled into full-scale data models 🡪 Universal / Industry-specific data models
* Reduced development time
* Higher model quality
* Information Engineering: A data-oriented methodology to create and maintain information systems
* Three-tiered client/web App/ database enterprise server database architecture
* Top-down planning–a generic IS planning methodology for obtaining a broad understanding of the IS needed by the entire organization
* Planning 🡪 align information technology with organization’s business strategies
* enterprise modeling and early conceptual data modeling

1. Identify strategic planning factors

* Organization goals
* Critical success factors
* Problem areas –weaknesses

1. Identify corporate planning objects

* Organizational units
* Organizational locations
* Business functions
* Entity types
* Information systems

1. Develop enterprise model

* Functional decomposition 🡪 break system description into finer
* Enterprise data model
* Planning matrixes 🡪 interrelationships between planning objects

1. Function-to-data entity
2. Location-to-function
3. Unit-to-function
4. IS-to-data entity

* Analysis 🡪 integrated conceptual data modeling / requirements analysis
* Logical Design 🡪 transactions, forms, information requirement structure, data integrity and security
* Physical design 🡪 develop technology and organizational specifications like DBMS
* Implementation 🡪 including coded programs, documentation, installation, training and conversion
* Maintenance 🡪 monitor, repair, enhance and periodic audits

Chapter 3

* Business Rules: Statements that define some aspect of the business / Control business behavior / Automated through DBMS software
* Declarative/what not how - Distinct - Business-oriented – Expressible – Precise/clear
* Good data name: Related to business / Meaningful / Unique / Readable
* Term – word or phrase with specific meaning
* Fact – association between two or more terms
* Relationship degrees specify number of entity types involved 🡪 unary / binary / ternary
* Relationship cardinalities specify how many of each entity type is allowed
* mandatory / optional 🡪 one / many
* Cardinality Constraints - the number of instances of one entity that can or must be associated with each instance of another entity
* Entity should be 🡪 important in system - multi value - multi attribute
* Entity should not be 🡪 A user / An output of the database system
* Attribute – property or characteristic of an entity or relationship type
* Required versus Optional Attributes
* Simple versus Composite Attribute
* Single-Valued versus Multivalued Attribute
* Stored versus Derived Attributes
* Identifier Attributes
* Associative Entity – combination of relationship and entity - a relationship with an attribute
* Ternary relationships and many-to-many relationships should be converted to associative entity
* Identifier (Key)–An attribute that uniquely identifies individual instances
* Will not change in value - Will not be null - No intelligent
* Candidate Identifier – an attribute that could be a key … satisfies the requirements for being an identifier
* Alternate Key: For example, in an employee table, empno is a primary key, empname is a alternate key that may not be unique but still helps in identifying a row of the table.
* Strong entities
* exist independently of other types of entities
* has its own unique identifier
* identifier underlined with single-line
* Weak entity
* dependent on a strong entity
* does not have a unique identifier
* Partial identifier underlined with double-line
* Entity box has double line
* Identifying relationship: links strong entities to weak entities

Chapter 4

* Sub-type: A subgrouping of the entities in an entity type that has attributes distinct from those in other subgroupings
* Super-type: A generic entity type that has a relationship with one or more subtypes
* An instance of a subtype is also an instance of the super-type
* Relationships at the super-type level indicate that all subtypes will participate in the relationship
* Generalization: The process of defining a more general entity type from a set of more specialized entity types. BOTTOM-UP
* Specialization: The process of defining one or more subtypes of the super-type and forming super-type/subtype relationships. TOP-DOWN
* Completeness Constraints: Whether an instance of a super-type must also be a member of at least one subtype
* Total Specialization Rule 🡪 double line 🡪 yes
* Partial Specialization Rule 🡪 single line 🡪 no
* Disjointness Constraints: Whether an instance of a super-type may simultaneously be a member of two (or more) subtypes
* Disjoint Rule 🡪 from only one subtype 🡪 d inside
* Overlap Rule 🡪 from more than one subtype 🡪 o inside
* Subtype Discriminator: An attribute of the super-type whose values determine the target subtype(s)
* Disjoint – a simple attribute 🡪 selects one sub-type
* Overlapping – a composite attribute 🡪 selects group of sub-types
* Entity cluster: Set of one or more entity types and associated relationships grouped into a single abstract entity type
* Classification of business rules:
* Derivation–rule derived from other knowledge
* Structural assertion–rule expressing static structure
* Action assertion–rule expressing controls of organizational actions
* Anchor Object – an object on which actions are limited
* Corresponding Objects – an object influencing the ability to perform an action on another business rule
* Types of action assertion
* Result from assertion 🡪 condition / integrity / authorization
* Form from assertion 🡪 enabler / timer / executive
* Rigor of the assertion 🡪 controlling / influencing

Chapter 5

* A relation is a named, two-dimensional table of data in 1st normal form
* For a table to be a relation
* Has unique name
* Each column unique name
* Each row is unique
* The order of columns and row is irrelevant
* Foreign keys: identifiers that enable a dependent relation to refer to its parent relation
* alternate key: when two attributes can be used as primary key and we can use any of them one of them is primary key and the other one is the alternate key the two are called candidate ID
* integrity constraints:
* domain constraint 🡪 set of values assigned to attribute 🡪 name , type , size
* entity integrity 🡪 must have primary key , not to be null
* referential integrity 🡪 any foreign key value must match a primary key value
* delete rules
* restrict 🡪 don’t delete parent side of dependent side
* cascade 🡪 delete dependent side with the parent side
* set-to-null 🡪 set foreign key to null when delete parent side

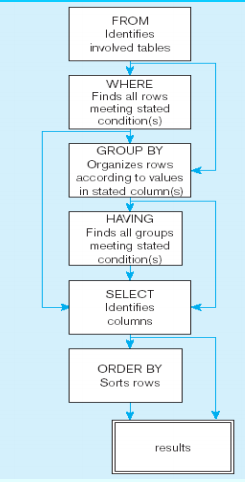
Chapter 7 - select

* Scalar aggregate: single value returned from SQL query with aggregate function
* Vector aggregate: multiple values returned from SQL query ith aggregate function (via GROUP BY)
* Database Schema: 🡪 start with abstract views then add details
* External Schema 🡪 User Views / subset of conceptual schema / repeats entity
* specific representation of an external view
* Easy to identify specific data required
* Facilitates designer’s job
* Ensures security
* Simplifies application development
* Conceptual Schema 🡪 E-R models
* Represents global view of the entire database
* All external views integrated into single global view
* Provides a relatively easily understood macro level view
* Independent of both software DBMS and hardware
* Internal model 🡪 Representation of the database as “seen” by the DBMS / SQL
* Maps the conceptual model to the DBMS / depends on software
* Logical independence: change internal model without affecting conceptual model
* Physical Schema 🡪 Physical structures
* Operates at lowest level of abstraction / the way data are saved on storage
* Physical independence: changes in physical model do not affect internal model

Chapter 7

* RDBMS: A database management system that manages data as a collection of tables in which all relationships are represented by common values in related tables
* Purpose of SQL 🡪 define data structures – standards and its enhancement – portability
* Benefits of standardization 🡪 reduce training cost – portability – productivity –longevity
* Catalog: A set of schemas that constitute the description of a database
* Schema: The structure that contains descriptions of objects created by a user
* Data Definition Language (DDL): Commands that define a database
* Data Manipulation Language (DML) Commands that maintain and query a database
* Data Control Language (DCL) Commands that control a database, including administering privileges and committing data
* Views provide users controlled access to tables
* Dynamic View🡪A “virtual table” created dynamically upon request - No data actually stored
* Materialized View🡪Copy or replication of data - Data actually stored – refreshed periodically
* Simplify query commands
* Enhance programming productivity
* Use little storage space
* May or may not be directly updateable – processing time is reference
* DBMS uses Indexes to Speed up random/sequential access to base table data

Chapter 8

* Sub query – placing an inner query (SELECT statement) inside an outer query
* Non-correlated – executed once for the entire outer query – doesn’t depend on outer
* Correlated – executed once for each row returned by the outer query – uses exits
* Best make sub-query in from statement
* Transaction = A discrete unit of work that must completely processed or not processed at all
* Functions – routines that return values and take input parameters
* Procedures – routines that do not return values and can take input or output parameters
* Triggers - Routines that execute in response to a database event

Data Normalization

* Primarily a tool to validate and improve a logical design so that it satisfies certain constraints that avoid unnecessary duplication of data
* A relation that contains minimal data redundancy and allows insertion , deletion without causing data inconsistencies
* Goal is to avoid anomalies:
* Insertion Anomaly –adding new rows forces user to create duplicate data
* Deletion Anomaly –deleting rows may cause a loss of data that would be needed for other future rows
* Modification Anomaly –changing data in a row forces changes to other rows because of duplication
* A table should not pertain to more than one entity type
* Functional Dependency: The value of one attribute (the determinant) determines the value of another attribute
* Why do these anomalies exist?
* Because there are multiple themes (entity types) in one relation. This results in duplication and an unnecessary dependency between the entities
* every non-key attribute is fully functionally dependent on the ENTIRE primary key no partial functional dependency
* transitive dependencies (functional dependencies on non-primary-key attributes)

