DBMS SUMMARY

Function of DBMS

* Data definition
* Data manipulation
* Data security and integrity
* Data recovery and STORAGE
* DATA DICTIONARY MAINTENANCE
* PERFORMANCE

Roles of a database administration

It is a person or group in charge of implementing DBMS in an organization. The DBA defines the schema and controls the 3 levels of the database.

RESPONSIBILITIES:

* Provides security and authorization
* Monitors recovery and backup
* Repairs damages caused by hardware or software failures
* Translates updates and queries in a nonprocedural language
* MAKES DECISION ON DATABASE STRUCTURE
* PLANS THE STORAGE STRUCTURE AND ACCESS STRATEGY
* Provides the support of the users
* Monitors performance and responds to change in requirements

DBMS CONSISTS OF 2 parts: DB and MS

Data -> Record -> Table -> Database

Facts -> data collection -> record collection -> table collection

Columns = field, attributes, domain \*all data here are of the same data type

Rows = tuples, records

DBMS is a collection of interrelated data and programs to view the data. The primary goal of DBMS is to provide a way to store and retrieve database info conveniently and efficiently. IT IS A COMPUTERIZED RECORD KEEPING SYSTEM.

Importance of management system is to create rules to maintain the database. the rules include:

* Attribute selection
* Attribute relation
* Row insertion and deletion
* Database integrity

DATA -> INFO -> KNOWLEDGE

89 -> MARKS: 89 -> PASSED

FILE SYSTEM

* Storing AND RETRIEVING DATA IS NOT EFFICIENT
* IT HAS NO CRASH RECOVERY MECHANISM
* POOR DATA PROTECTION
* Data redundancy and inconsistency
* Difficult data access
* Failure of the computer system

History

* 1950: magnetic tape, punch cards, printers
* 1960: hard disks
* 1970: relational model
* 1980: network and hierarchical
* 1990: SQL languages, WWW

Evolution

* FMS: single large file
* Hierarchical DB: PARENT CHILD RELATIONSHIP
* Network DB: USED POINTERS TO SHOW RELATIONSHIP
* Relational DB: USE OF TABLES

PROS AND CONS

* REDUNDANCY CONTROL
* Data SHARING AND INTEGRITY
* DATA SECURITY AND CONSISTENCY
* DATA ACCESS AND INDEPENDENCE
* ENFORCED STANDARDS
* COMPLEX
* HIGH MEMORY USAGE
* FAILURE IMPACTS THE ENTIRE USERS
* SLOW

DATA ABSTRACTION:

* PHYSICAL (INTERNAL) VIEW/ SCHEMA: HOW THE DATA IS ACTUALLY Stored. it covers the file structure and file organization used to store the data
* LOGICAL (CONCEPTUAL): WHAT DATA IS STORED AND THE RELATIONSHIPS AMONGST THEM. INTEGRATES PHYSICAL DATA Independence. It represents all entities, attributes and their relationships, the constraints and the security and integrity of information. It has 2 types:
* Logical: it indicates that the conceptual schema can be changed without affecting the existing external schema
* Physical: indicates that the physical storage structure can be changed without affecting the conceptual schema
* View (external): describes part of the db. Can only view the part that concerns you.it encapsulates users from the details of the other views.

Instances and schema: the collection of information stored at a particular moment is called an instance. The overall design is called the schema.it is a description of a particular collection of data using data models. Physical schema, logic schema, subschemas.

Data models: a collection of conceptual tools for describing data, their relationships, consistency constraints and semantics

Categories

* Relational: uses tables to represent the data and its relationships among said data. It is a record based model.
* Entity-relational: uses entities and relationships. An entity is an object that is distinguishable from other objects.
* Object based: E-R plus encapsulation, methods, object identity
* Semi structured: permits specification of data where individual data items have different set of attributes. Extensible markup language is used to represent this data

Database languages

* Data definition language to identify the schema. It also specifies additional properties of a data. Has commands
* Create, alter, rename, DROP, COMMENT, TRUNCATE
* Data manipulation language to express database queries and updates. HAS COMMANDS: INSERT, DELETE, SORT, UPDATE

Together they make up the SQL language.

1. Data manipulation language: enables users to manipulate data in a given mode. Access types are:

* Retrieval, Insertion, Deletion, modification

Types of DML are:

* Procedural specify what data is needed and how to get it
* Declarative specify what data is needed without specifying how to get the data. Also called nonprocedural.

Query is a statement that request the retrieval of information. The DML portion for information retrieval is query language.

1. Data definition language: specify the storage structure and access method. use data storage and definition language where the data values satisfy certain consistency constraints.
2. DATA CONTROL LANGUAGE: GRANT, REVOKE, DENY
3. TRANSACTION CONTROL LANGUAGE: MANAGES CHANGES MADE BY DML STATEMENTS. Commands are: commit, save point, rollback, set transaction.

* Domain constraints: a domain of values to every attribute(field), each value in a field must correspond to its data type.
* Referential integrity: a value in a relation also appears as a value in another relation set.
* Assertion: conditions the database must always satisfy. As long as assertions are not violated, modifications can be made.
* Authorization: read, insert, update, delete. In an all or some variation

Output is stored in data dictionary which has data about data. It is only accessible by the database itself.

Types of data dictionary

* Integrated: included with the dbms, limited metadata to the DBMS
* Stand-alone: not inbuilt in the dictionary, is more flexible.

Personalities of Data dictionaries

* Active: automatically updated
* Passive: Not updated automatically and usually requires a batch process to run.

Database users and interface –Types

* Naïve end users: users who interact with the system by invoking programs.
* Application programmers: write apps. Rapid app dev tools are tools that allow a programmer to construct forms and reports without a program.
* Sophisticated users: interact with the system without writing programs. They create query requests and send it to the query processor which breaks down statements into instructions understood by the storage manager. They perform tasks such as
* Online analysis processing: views summary of data in different ways.
* Data mining: allows them to find certain patter in data
* Specialized users write specialized database apps that do not fit into traditional data processing frameworks

Database architecture: influenced by the underlying computer system. The functional component is divided into.

1. Storage manager.
2. Query inquiry processor

2-tier apps are partitioned into 2 parts: uses query languages extensively

3-tier apps act as a front end and doesn’t call any direct database call. They are more appropriate for large apps.

Query processor:

* DDL interpreter: interprets DDL statement and records them in data dictionary
* DMl Compiler: translates query language into an evaluation plan consisting of low level instructions the query evaluation engine understands
* Query optimization: picking the lowest cost
* Query evaluation engine executes low-level instructions from the DML compiler
* Storage manager: provides an interface between the low level data and the application programs and queries submitted to the system (it is responsible for the interaction with the file manager). Its components are:
* Authorization and integrity manager: tests the satisfaction of integrity constraints and check authority if users to data
* Transaction manager: ensure the database remains consistent despite failures
* File manager: manages space allocation and data structure in the disk
* Buffer manager: fetches data from disk to memory handles data sizes larger than the memory.
* Transaction manager: a transaction is a collection of operations that performs a single logical function in a database application and each transaction should have atomicity and consistency.
* Conceptual database design: database design techniques
* ER modeling (Top down approach)
* Normalization (bottom top approach)

1. ER modeling: a graphical technique for understanding and organizing independent data.

* ENTITY: any independent existence that collects data.
* Entity instance: a particular member of an entity type. An entity with its own key attribute is a regular entity
* Weak entity: an entity that depends on another entity for its existence and doesn’t have its own key attribute
* Attribute: characteristics of entities.
* Domain of attributes: set of values an attribute can take.
* Key attribute: unique attribute usually the ID. If it consists of multiple attributes it becomes a composite key.
* Simple attribute: cannot be divided into simpler components
* Composite attributes: can be simplified
* Single valued attribute: takes only one value for each entity instance
* Multi value attribute: takes multiple values
* Stored attributes: attributes to be stored permanently
* Derived attribute: calculated or derived from other attributes
* Relationships: associations between entities
* Identifying relationship: between weak and strong entities
* Degree of relationship: number of entities involved in the relationship. Unary, binary, ternary -> 1, 2, >2
* Cardinality: how many of each entity is involved 1-1, 1-n, m-1, M-n.

Relationship participation

* Total: all entity instances are connected to all entities. Also called existence dependency
* Partial: not all instances are connected to the entity

Advantages and disadvantages of ER model

* Simple and easy to understand
* helps in physical data creation
* Can be generalized or specialized based on need
* Helps in database design
* Gives a higher level description of the system
* Physical designs may contain ambiguity or inconsistencies
* can lead to misinterpretation

relational model: consists of a collection of tables which each have a unique name. the tables represent relationships

relation -> tuple -> attribute

table -> row -> column

database schema: logical design of a databases

database instance: snapshot of the data in a database

super key: a set of attribute that allows us to uniquely identify a tuple in a table it is the primary key. If a relation includes the primary key of another relation it is called the foreign key.

Structured Query language:

SELECT \* FROM Customers; selects all rows from customers

UPDATE (MUST HAVE WHERE), DELETE MUST HAVE WHERE), INSERT INTO (ADDS BY ROW), CREATE DATABASE, ALTER DATABASE, CREATE TABLE, ALTER TABLE, DROP TABLE, CREATE INDEX, DROP INDEX, MIN, MAX, COUNT, AVG, SUM

ALTER = UPDATE, DROP = DELETE, WHERE = FILTER RECORD BASED ON CONDITION, ORDER BY = SORT (USE DESC OR ASC)

COUNT = NUMBER OF ROWS THAT MEET CONDITION

Meta data of data dictionary: active and passive

Components of DBMS

* data: most important, constitutes the database
* procedures: instruction and rules that govern the design and use of the database
* software: lies between the physical database and the user
* users: people involved in the access and retrieval of data in the database.

disadvantages of DBMS

* complexity
* size
* cost
* performance
* vulnerability of system

indexed sequential access method

architecture of DBMS

* external level
* conceptual
* internal

objective of 3-level architecture: to separate each user view of the data from the way the database id physically represented. The internal structure should be unaffected while making changes to the physical aspects of storage. The DBA should be able to change the conceptual structure of the database without affecting the other users

external schema; different view of data can have more than 1

conceptual schema: describes the entities, attributes and their relationships with the integrity constraints. Can have only one

internal schema: complete description of the internal model. Can have only one

external to conceptual mapping

conceptual to internal mapping

difference between file and database management

* master file: they remain static
* transaction file: files which are dynamic in nature, they can make changes.
* Instances: the situation of the database at a particular moment in time
* subschema: inherits the same properties as a schema

limitations of file processing system

* separated and isolated data
* difficulty in data representation

information retrieval is a process of accessing data resources. There are 3 types

* Boolean
* Vector
* Probabilistic

system is a software that provides access to sources of information. 2 types

* Index system,
* Query system

Methods of information retrieval

* Recall
* Recognition
* Renaming

Difference between data and information retrieval system

* Information retrieval is concerned with free and unstructured data while data retrieval is concerned with structured and relations of a database

Five information retrieval tools

* Search engine
* Internet
* Email
* Opac: online public access catalog

Major components of IRS

* Storage
* Searching
* Natural language processing
* Ranking and relevance storage
* Query processing

Benefits of information retrieval

* Enables users to quickly access relevant information

DEPARTMENTS

Employee

Project