

Introduction to Databases

Madiba Hudson-Quansah

CONTENTS

CHAPTER 1	INTRODUCTION	PAGE 2
1.1	Uses of Databases	2
1.2	File Based Systems	2
	Disadvantages of File-Based Systems — 2	
1.3	Database Management Systems (DBMS)	3
	Database — 3 • Database Management System (DBMS) — 3	
	1.3.2.1 Properties of a DBMS	3
	1.3.2.2 Difference between File-Based Systems and DBMS	4
	1.3.2.2.1 The Self-Describing Nature of a Database System	4
	1.3.2.2.2 Multiple Views of Data	4
	1.3.2.2.3 The Concurrent Access to the Data	4
1.4	(Database) Application Programs	4
	Views — 4	
1.5	Comparison of File-Based Systems and DBMS	4
	Advantages of Using the DBMS Approach — 4 • Disadvantages of Using the DBMS Approach — 5	
1.6	Components of a Database System	5
CHAPTER 2	DATABASE SYSTEM ENVIRONMENT	PAGE 6
	Database Users / Roles — 6	
2.1	ANSI-SPARC Three-Level Architecture	6
	External Level — 7 • Conceptual Level — 7 • Internal Level — 7 • Data Independence — 7	
	2.1.4.1 Logical Data Independence	7
	2.1.4.2 Physical Data Independence	8
2.2	Data Models	8
2.3	Database Languages	8
	Data Definition Language (DDL) — 8 • Data Manipulation Language (DML) — 8	
2.4	The Relational Model	9
	Terminology — 9	
	2.4.1.1 Database	9
	2.4.1.2 Relation	9
	2.4.1.3 Attribute	9
	2.4.1.4 Tuple	9
	2.4.1.5 Type	9
	2.4.1.6 Schema	9
	2.4.1.7 Instance	9
	2.4.1.8 Null values	9
	2.4.1.9 Key	10
2.5	Software Architecture	10

Chapter 1

Introduction

1.1 Uses of Databases

Retail - Department stores, supermarkets, mail order companies, etc.

Social Media - Store information about users, their friends, and their activities.

1.2 File Based Systems

Definition 1.2.1: File-Based Approach

A collection of application programs that perform services for the end-users, where each program defines and manages its own data

1.2.1 Disadvantages of File-Based Systems

- Space intensive
- Data redundancy / Duplication - Due to each program having and controlling its own data, if data is needed by multiple programs, it must be duplicated.
- Separation and isolation of data - When data is scattered across different files, it becomes difficult to access data that should be available.
- Data Dependence / Program-Data Dependence - The inherent relationship between the application program and the data it uses. This leads to:
 - Difficulty in updating data
- Incompatible file formats.
- Fixed Queries / Proliferation of application programs - Due to the nature of file-based systems creating new queries is difficult as a developer would have to write the new queries into the program.
- Difficulty in accessing data
- Inconsistent data

1.3 Database Management Systems (DBMS)

1.3.1 Database

Definition 1.3.1: Database

A *shared* collection of *logically* related data, and descriptions of this data, designed to meet the information needs of an organization. / A set of named relations.

Definition 1.3.2: System Catalogue / Data Dictionary

A description of the data in the database (metadata).

Definition 1.3.3: Entity / Record

A distinct real world object that is to be represented in a database / Row.

Definition 1.3.4: Attribute

A property that describes an entity. / Column.

Definition 1.3.5: Relationship

An association between entities. / Table

1.3.2 Database Management System (DBMS)

Definition 1.3.6: Database Management System (DBMS)

A software system that enables users to define, create, maintain, and control access to the database. A DBMS usually provides the following functions:

Data Definition Language (DDL) - Allows users to define a database by describing the data types, structures, and constraints on the data to be stored in the database.

Data Manipulation Language (DML) - Allows users to insert, update, delete, and retrieve data from the database. The DML, then provides a **query language**, which is used for inquiry and reporting.

Controlled Access -

- Ensures that only authorized users can access the database,
- Ensures the consistence of the stored data is maintained
- Manages concurrent access to the database
- Provides a recovery system to ensure consistency of data in the presence of system failures.

1.3.2.1 Properties of a DBMS

Massive Be able to manage large amounts of data.

Persistence Data should be available even after the application has been closed.

Safe Hardware and software failures should not cause data loss.

Multi-user / Concurrent Multiple users should be able to access the database at the same time>

Convenient Be able to declaratively specify queries and operations.

Efficient Be able to perform operations and process queries quickly.

Reliable Be able to recover from failures.

1.3.2.2 Difference between File-Based Systems and DBMS

1.3.2.2.1 The Self-Describing Nature of a Database System

The database system contains the complete definition of the database structure and constraints. This definition is

1.3.2.2.2 Multiple Views of Data

Each user may see a different view of the database, and the DBMS must control the access of these users to the database.

1.3.2.2.3 The Concurrent Access to the Data

Each user must be able to access the data concurrently, and the DBMS must include concurrency control techniques to ensure that the user's operations are correctly synchronized.

Definition 1.3.7: Database System

A DBMS together with the database itself.

1.4 (Database) Application Programs

Definition 1.4.1: (Database) Application Programs

Programs that interact with the DBMS to access the database, using the DBMS's DML and query language.

1.4.1 Views

Definition 1.4.2: View / View Mechanism

A subset of the database, defined using queries. Views provide an abstracted view of the database, hiding irrelevant information from the end-user.

Views not only serve as an abstraction layer but also provide:

A Level of Security - Views can be used to exclude sensitive data to unauthorized parties.

Customization - Views can be used to better present data for end-users

Consistency - Views can be used to preserve the state of the database even though the underlying data may have changed.

1.5 Comparison of File-Based Systems and DBMS

1.5.1 Advantages of Using the DBMS Approach

- Control of data redundancy - Data is stored in a central location, and is not duplicated unnecessarily across multiple programs. As some data will need to be duplicated for various reasons such as performance.
- Data Consistency - If a data item is stored only once in the database, an update to it will reflect in all places it is used, ensuring consistency.
- Data Sharing - Data can be shared across departments, and applications without the need for duplication.
- Improved Data Integrity - Database integrity refers to the validity and consistency of stored data. Integrity is defined in terms of constraints, which are consistency rules that the database is not allowed to violate. With DBMS, the Database Admin can define constraints that the DBMS will enforce.

- Economy of Scale - The cost of developing and maintaining the database is spread over all the applications that use the database.

1.5.2 Disadvantages of Using the DBMS Approach

- Complexity - The DBMS is a complex piece of software, and as such, it requires a high level of expertise to manage.
- Size - The complexity and breadth of functionality usually makes the DBMS large and resource-intensive.
- Cost of DBMS - The cost of acquiring a DBMS can be high.
- Additional hardware costs - The disk storage requirements of a DBMS and the database may require the purchase of additional storage space.
- Higher impact of a failure - The centralization of resources increases the vulnerability of the system to failure as a failure will affect all the applications that use the database which due to centralization, is a lot.
- Performance - The DBMS may not be as efficient as a custom-built file-based system.

1.6 Components of a Database System

- Users - Application programs, DBA, End-users
- Software - Controls the organization, storage, management and retrieval of data.
- Hardware - The physical devices used to store and process the data.
- Data - The data to be stored in the database.

Chapter 2

Database System Environment

2.0.1 Database Users / Roles

Database Administrators Responsible for authorizing access to the database for coordinating and monitoring its use. Acquiring software and hardware resources, controlling its use, and monitoring efficiency of operations.

Data Administrators Responsible for the management of the organization's data resources, including the database planning, development, maintenance of standards, policies, and procedures.

Database Designers Responsible for defining the content structure, constraints, and functions or transactions of the database. Database designers can be split into two categories:

- Logical Database Designers - Focus on the logical structure of the database, i.e. the constraints on data to be stored in the database (business rules).
- Physical Database Designers - Focuses on how the logical database is to be actually implemented. This involves:
 - Mapping the logical structure to a set of tables and integrity constraints
 - Selecting storage structures and access methods
 - Designing security measures required on the data.

Application Developers Responsible for developing the application programs that provide required functionality for the end users, using the implemented database.

End Users Clients of the database, which has been designed and implemented to serve their information needs. They can be classified according to the way they interact with the system.

- Naive Users - Unsophisticated users who interact with the system through application programs.
- Casual - Users who access the database occasionally, and may not be familiar with the system.
- Sophisticated - Users who interact with the system regularly and are familiar with the structure of the database and the facilities provided by the DBMS.
- Stand-alone - Users who maintain personal databases using ready-to-use packaged applications.

2.1 ANSI-SPARC Three-Level Architecture

Definition 2.1.1: ANSI-SPARC Three-Level Architecture

American National Standards Institute (ANSI) Standards Planning and Requirements Committee (SPARC). A database architecture that separates the user's view of the database from the way the data is physically stored.

This architecture is based on the idea that a database system should be divided into three levels of abstraction:

- External Level
- Conceptual Level
- Internal Level

2.1.1 External Level

Definition 2.1.2: External Level

The user's view of the database. Describes the part of the database that is relevant to the user.

This level consists of a number of external views of the database. Each external view includes only the entities, attributes and relationships that are relevant to a particular user group.

2.1.2 Conceptual Level

Definition 2.1.3: Conceptual Level

The community view of the database. Describes what data is stored in the database and the relationships between the data.

This level contains the logical structure of the database as seen by the Database Administrator, i.e., the complete view of the database without considering implementation details. The conceptual level represents:

- All entities, attributes, and relationships
- Data integrity constraints
- Semantic information about the data
- Security information.

»

2.1.3 Internal Level

Definition 2.1.4: Internal Level

The physical representation of the database on the computer. Describes how the data is stored in the database.

- Storage space allocation of data and indexes
- Record description for storage
- Record placement (pointers)
- Data compression and encryption techniques

2.1.4 Data Independence

2.1.4.1 Logical Data Independence

Definition 2.1.5: Logical Data Independence

Possibility for addition/removal of new entities, attributes or relationships. Indicates that the conceptual schema can be changed without affecting the existing external schemas.

This mainly affects the conceptual and external levels of the database, enabling a form of data abstraction from the external views and the conceptual structure of the database.

2.1.4.2 Physical Data Independence

Definition 2.1.6: Physical Data Independence

Possibility for changes to storage structures. Indicates that the physical storage structures or devices could be changed without affecting conceptual schema.

This mainly affects the conceptual and internal levels of the database, decoupling the conceptual description of the database from the actual physical implementation of the database.

2.2 Data Models

Definition 2.2.1: Data Model

An integrated collection of concepts for describing and manipulation of data. The data model provides the necessary means to achieve data abstraction, which is the process of hiding the low-level details of the database and showing only the relevant data to the user.

- Object-Based
- Record-Based
- Physical Data Model

2.3 Database Languages

Definition 2.3.1: Database Languages

A language that enables the user to create and maintain the database and provide ways to access and manipulate the data stored in the database.

- Data Definition Language (DDL)
- Data Manipulation Language (DML)

2.3.1 Data Definition Language (DDL)

Definition 2.3.2: Data Definition Language

A language used to define and name, entities, their attributes, and relationships required for the application, along with integrity constraints.

2.3.2 Data Manipulation Language (DML)

Definition 2.3.3: Data Manipulation Language

2.4 The Relational Model

2.4.1 Terminology

2.4.1.1 Database

Definition 2.4.1: Database

A set of named relations.

2.4.1.2 Relation

Definition 2.4.2: Relation / Table

A table in a relational database.

2.4.1.3 Attribute

Definition 2.4.3: Attribute

A column in a table. Each relation has a named set of attributes.

2.4.1.4 Tuple

Definition 2.4.4: Tuple

A row in a table. Each tuple has a value for each attribute.

2.4.1.5 Type

Definition 2.4.5: Type / Domain

A description of the types of operations and values an attribute can have.

2.4.1.6 Schema

Definition 2.4.6: Schema

A structural description of relations in a database / A structural description of the tables in a database. A database differs from a database by the fact that a database is the implementation of the schema.

2.4.1.7 Instance

Definition 2.4.7: Instance

The actual data in a database at a particular point in time.

2.4.1.8 Null values

Definition 2.4.8: Null

Unknown / Undefined

2.4.1.9 Key

Definition 2.4.9: Key

An attribute unique to each tuple, used to differentiate entities / tuples.

2.5 Software Architecture

Definition 2.5.1: Presentation Layer

Medium of access.

Definition 2.5.2: Application Layer

Software that is accessed by the presentation layer.

Definition 2.5.3: Database Layer

The database

Note:-

Machine means server.

- One-Tier architecture - Presentation layer , Application layer , Database layer on the same machine.
- Two-Tier architecture - Database tier (Application and Database layers) on one machine and client tier (Presentation layer) on another.
- Three-Tier architecture - Each layer on its own machine
- N-Tier architecture - Multiple instances of tiers using replication procedures and load balancing.