Chapter 1: Introduction

Database System Concepts, 6th Ed.

Outline

- The Need for Databases
- Key Features
- DDL, DML
- Relational Databases
- Database Design
- Storage Manager
- Query Processing
- Transaction Manager
- File System and Database

Database Management System (DBMS)

- DBMS contains information about a particular enterprise
  - Collection of interrelated data
  - Set of programs to access the data
  - An environment that is both *convenient* and *efficient* to use
- Database Applications:
  - Banking: transactions
  - Airlines: reservations, schedules
  - Universities: registration, grades
  - Sales: customers, products, purchases
  - Online retailers: order tracking, customized recommendations
  - Manufacturing: production, inventory, orders, supply chain
  - Human resources: employee records, salaries, tax deductions
- Databases can be very large.
- Databases touch all aspects of our lives

University Database Example

- Application program examples
  - Add new students, instructors, and courses
  - Register students for courses, and generate class rosters
  - Assign grades to students, compute grade point averages (GPA) and generate transcripts
- In the early days, database applications were built directly on top of file systems

Drawbacks of using file systems to store data

- Data redundancy and inconsistency
  - Multiple file formats, duplication of information in different files
- Difficulty in accessing data
  - Need to write a new program to carry out each new task
- Data isolation
  - Multiple files and formats
- Integrity problems
  - Integrity constraints (e.g., account balance > 0) become “buried” in program code rather than being stated explicitly
  - Hard to add new constraints or change existing ones

Drawbacks of using file systems to store data (Cont.)

- Atomicity of updates
  - Failures may leave database in an inconsistent state with partial updates carried out
  - Example: Transfer of funds from one account to another should either complete or not happen at all
- Concurrent access by multiple users
  - Concurrent access needed for performance
  - Uncontrolled concurrent accesses can lead to inconsistencies
    - Example: Two people reading a balance (say 100) and updating it by withdrawing money (say 50 each) at the same time
- Security problems
  - Hard to provide user access to some, but not all, data

Database systems offer solutions to all the above problems

Levels of Abstraction

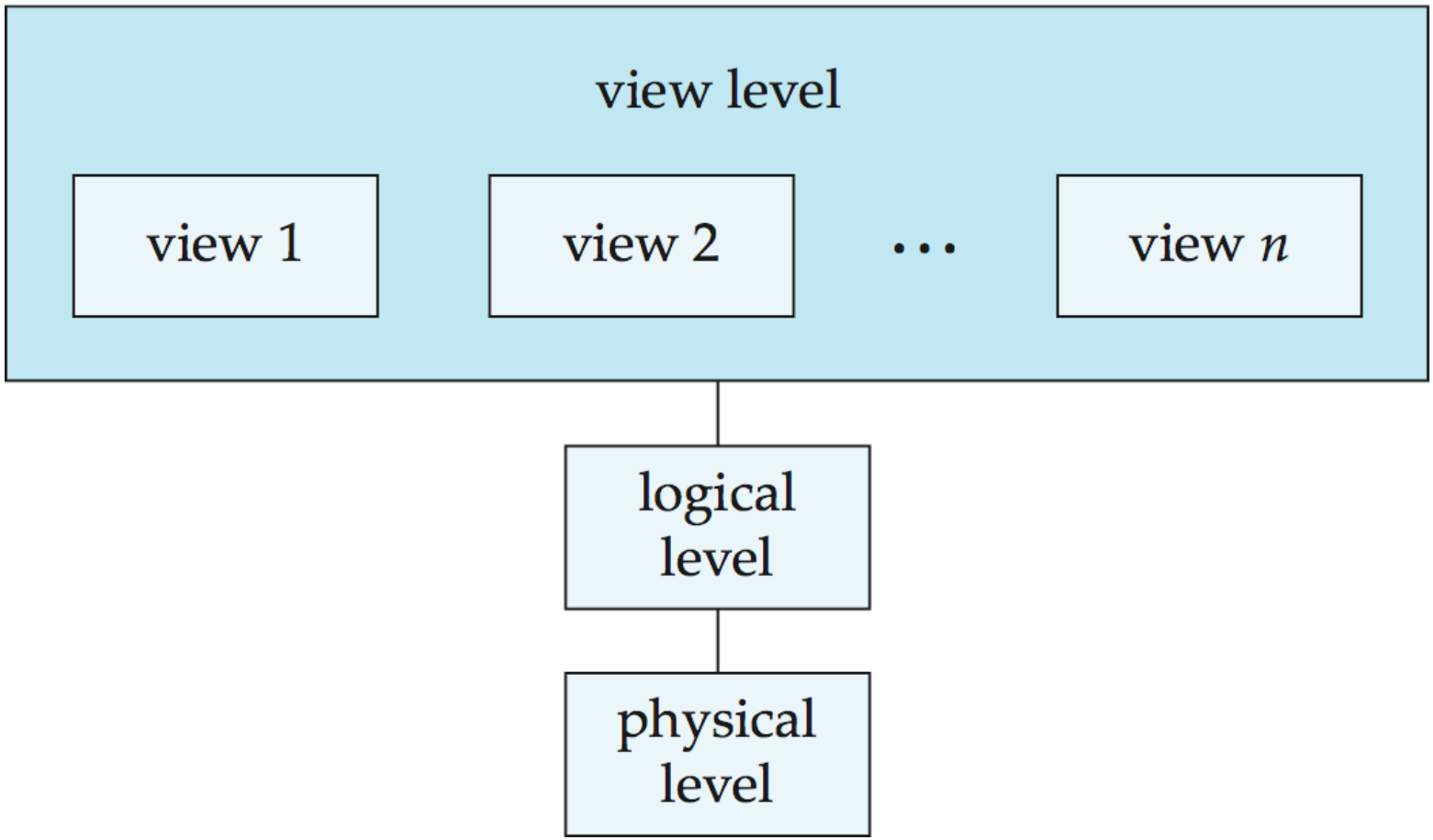
- **Physical level:** describes how a record (e.g., instructor) is stored.
- **Logical level:** describes data stored in database, and the relationships among the data.

```
type instructor = record
        ID : string;
        name : string;
        dept_name : string;
        salary : integer;
    end;
```

- **View level:** application programs hide details of data types. Views can also hide information (such as an employee's salary) for security purposes.

View of Data

An architecture for a database system



Database Components

- Field
- Record
- Value (data)

Field = Attribute = Column
Record = Tuples = Row

- Field can contain same value. Do not provide proper information.
- Record can contain different values. Find complete information.

Roll	Name	Age	Gender
1010	A	23	M
1011	B	25	F

Database Management System

- Database Management System is a software or technology used to manage data from a database.
- Some popular databases are MySQL, Oracle, MongoDB, etc.
- DBMS provides many operations e.g. creating a database, Storing in the database, updating an existing database, delete from the database.
- DBMS is a system that enables you to store, modify and retrieve data in an organized way.
- It also provides security to the database.

Key Features of DBMS

Data modeling: A DBMS provides tools for creating and modifying data models, which define the structure and relationships of the data in a database.

Data storage and retrieval: A DBMS is responsible for storing and retrieving data from the database, and can provide various methods for searching and querying the data.

Concurrency control: A DBMS provides mechanisms for controlling concurrent access to the database, to ensure that multiple users can access the data without conflicting with each other.

Data integrity and security: A DBMS provides tools for enforcing data integrity and security constraints, such as constraints on the values of data and access controls that restrict who can access the data.

Backup and recovery: A DBMS provides mechanisms for backing up and recovering the data in the event of a system failure.

DBMS can be classified into two types: Relational Database Management System (RDBMS) and Non-Relational Database Management System (NoSQL or Non-SQL)

RDBMS: Data is organized in the form of tables and each table has a set of rows and columns. The data are related to each other through primary and foreign keys.

NoSQL: Data is organized in the form of key-value pairs, documents, graphs, or column-based. These are designed to handle large-scale, high-performance scenarios.

Data Definition Language

DDL is the short name for Data Definition Language, which deals with database schemas and descriptions, of how the data should reside in the database.

CREATE: to create a database and its objects like (table, index, views, store procedure, function, and triggers)

ALTER: alters the structure of the existing database

DROP: delete objects from the database

TRUNCATE: remove all records from a table, including all spaces allocated for the records are removed

COMMENT: add comments to the data dictionary

RENAME: rename an object

Data Manipulation Language

DML is the short name for Data Manipulation Language which deals with data manipulation and includes most common SQL statements such SELECT, INSERT, UPDATE, DELETE, etc., and it is used to store, modify, retrieve, delete and update data in a database.

SELECT: retrieve data from a database

INSERT: insert data into a table

UPDATE: updates existing data within a table

DELETE: Delete all records from a database table

MERGE: UPSERT operation (insert or update)

CALL: call a PL/SQL or Java subprogram

EXPLAIN PLAN: interpretation of the data access path

LOCK TABLE: concurrency Control

Data Control Language

DCL is short for Data Control Language which acts as an access specifier to the database.(basically to grant and revoke permissions to users in the database

GRANT: grant permissions to the user for running DML(SELECT, INSERT, DELETE,…) commands on the table

REVOKE: revoke permissions to the user for running DML(SELECT, INSERT, DELETE,…) command on the specified table

Transactional Control Language

TCL is short for Transactional Control Language which acts as an manager for all types of transactional data and all transactions. Some of the command of TCL are

Roll Back: Used to cancel or Undo changes made in the database

Commit: It is used to apply or save changes in the database

Save Point: It is used to save the data on the temporary basis in the database

Advantages of DBMS

Data organization: A DBMS allows for the organization and storage of data in a structured manner, making it easy to retrieve and query the data as needed.

Data integrity: A DBMS provides mechanisms for enforcing data integrity constraints, such as constraints on the values of data and access controls that restrict who can access the data.

Concurrent access: A DBMS provides mechanisms for controlling concurrent access to the database, to ensure that multiple users can access the data without conflicting with each other.

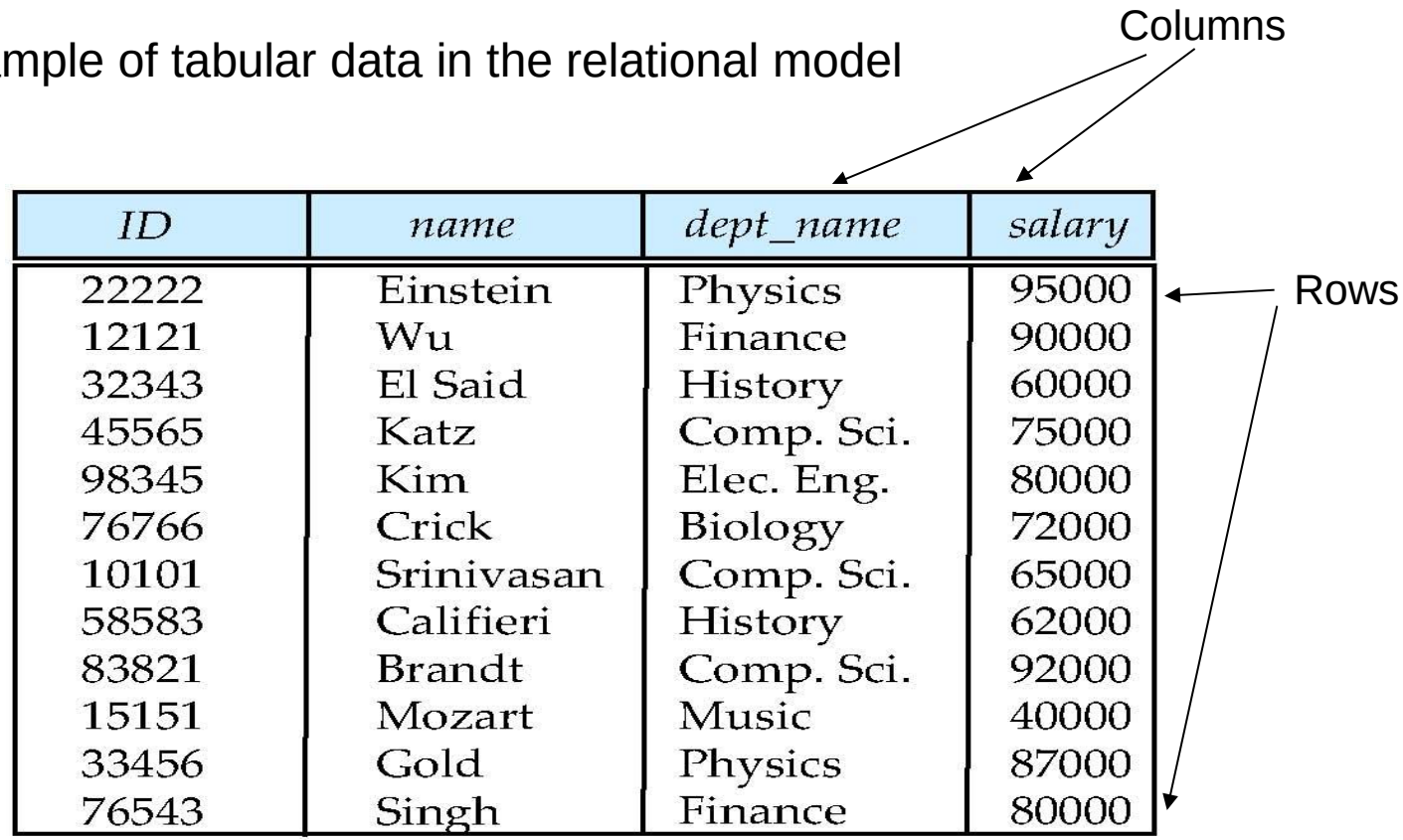
Data security: A DBMS provides tools for managing the security of the data, such as controlling access to the data and encrypting sensitive data.

Backup and recovery: A DBMS provides mechanisms for backing up and recovering the data in the event of a system failure.

Data sharing: A DBMS allows multiple users to access and share the same data, which can be useful in a collaborative work environment.

Relational Model

- All the data is stored in various tables.
- Example of tabular data in the relational model



ID	*name*	*dept_name*	*salary*
22222	Einstein	Physics	95000
12121	Wu	Finance	90000
32343	El Said	History	60000
45565	Katz	Comp. Sci.	75000
98345	Kim	Elec. Eng.	80000
76766	Crick	Biology	72000
10101	Srinivasan	Comp. Sci.	65000
58583	Califieri	History	62000
83821	Brandt	Comp. Sci.	92000
15151	Mozart	Music	40000
33456	Gold	Physics	87000
76543	Singh	Finance	80000

(a) The *instructor* table

A Sample Relational Database

ID	*name*	*dept_name*	*salary*
22222	Einstein	Physics	95000
12121	Wu	Finance	90000
32343	El Said	History	60000
45565	Katz	Comp. Sci.	75000
98345	Kim	Elec. Eng.	80000
76766	Crick	Biology	72000
10101	Srinivasan	Comp. Sci.	65000
58583	Califieri	History	62000
83821	Brandt	Comp. Sci.	92000
15151	Mozart	Music	40000
33456	Gold	Physics	87000
76543	Singh	Finance	80000

(a) The *instructor* table

dept_name	*building*	*budget*
Comp. Sci.	Taylor	100000
Biology	Watson	90000
Elec. Eng.	Taylor	85000
Music	Packard	80000
Finance	Painter	120000
History	Painter	50000
Physics	Watson	70000

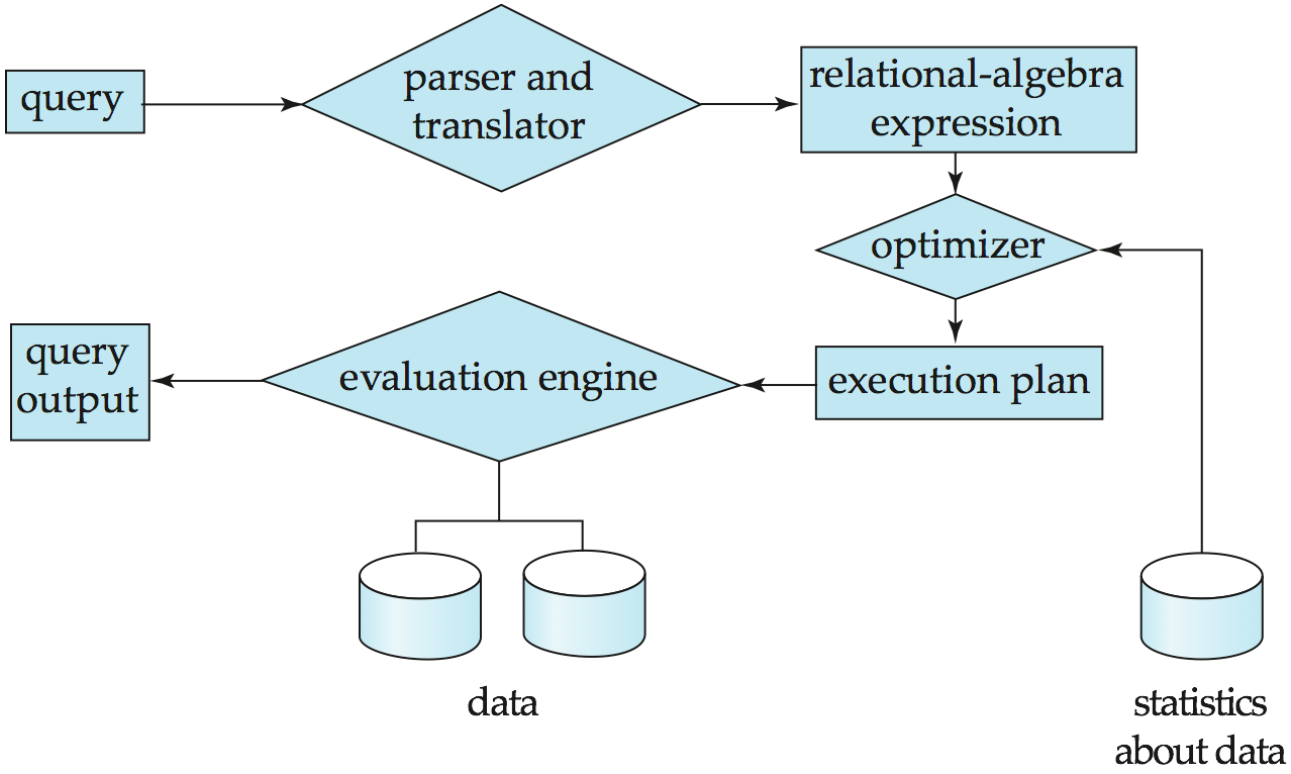
(b) The *department* table

SQL

- The most widely used commercial language
- SQL is NOT a Turing machine equivalent language
- SQL is NOT a Turing machine equivalent language
- To be able to compute complex functions SQL is usually embedded in some higher-level language
- Application programs generally access databases through one of
  - Language extensions to allow embedded SQL
  - Application program interface (e.g., ODBC/JDBC) which allow SQL queries to be sent to a database

Query Processing

1. Parsing and translation
2. Optimization
3. Evaluation



Query Processing (Cont.)

- A parser is a software component that breaks down input data into smaller parts so that other software can understand it.

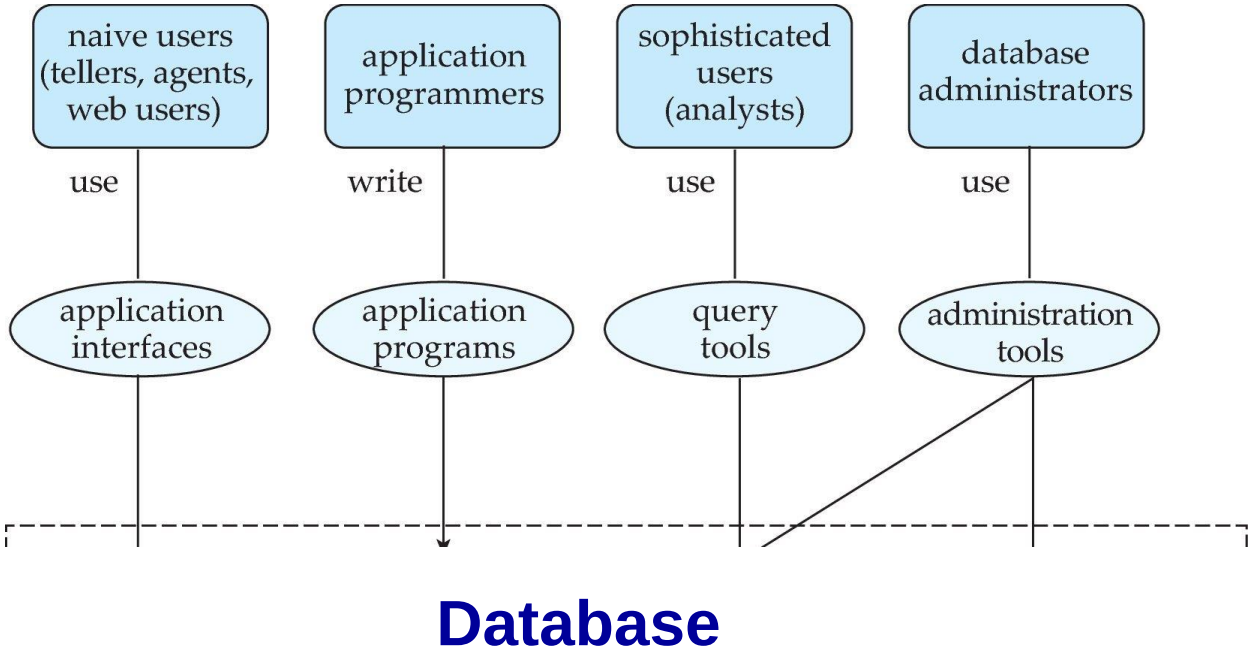
- Relational algebra consists of a certain set of rules or operations.

1. Selection(σ)
2. Projection(π)
3. 3. Union(U)
4. Set Difference(-)
5. Set Intersection(∩)
6. Cross Product(X)

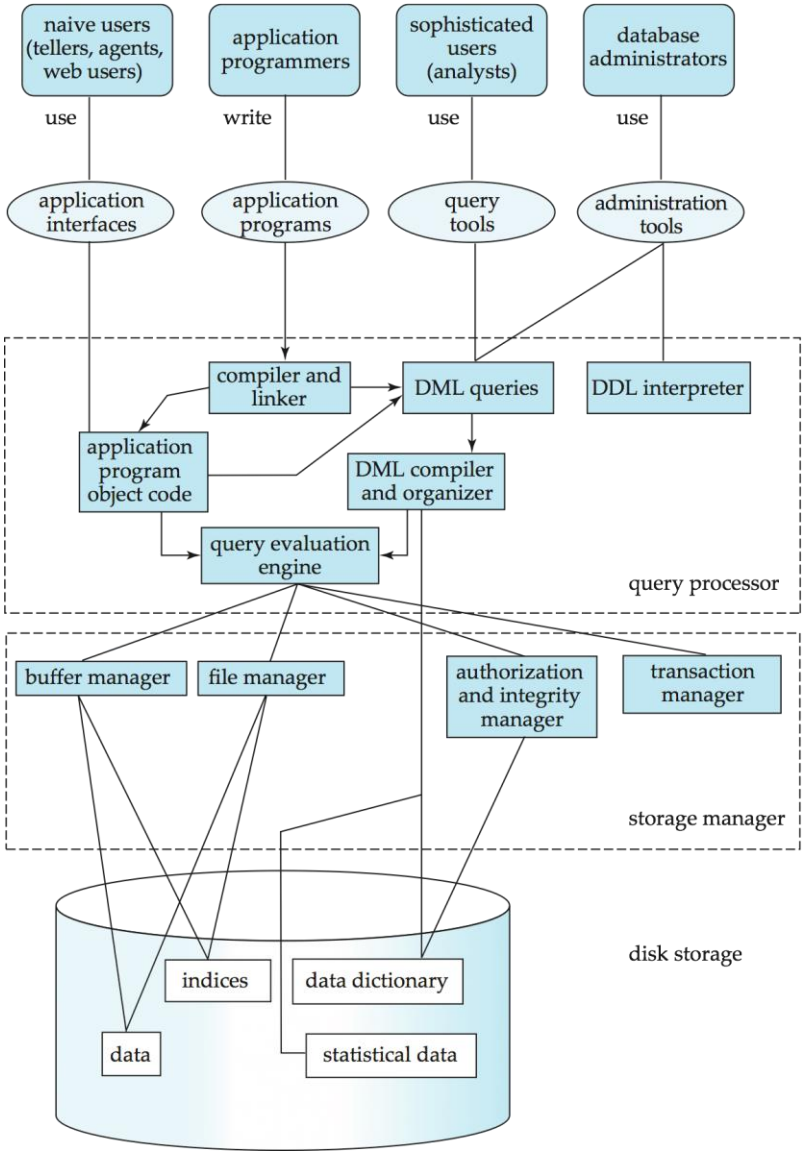
Transaction Management

- What if the system fails?
- What if more than one user is concurrently updating the same data?
- A **transaction** is a collection of operations that performs a single logical function in a database application
- **Transaction-management component** ensures that the database remains in a consistent (correct) state despite system failures (e.g., power failures and operating system crashes) and transaction failures.
- **Concurrency-control manager** controls the interaction among the concurrent transactions, to ensure the consistency of the database.

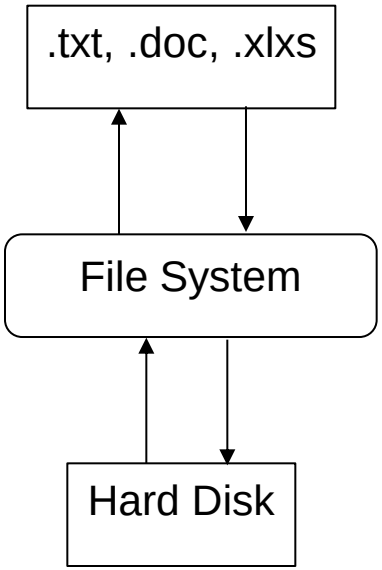
Database Users and Administrators



Database System Internals

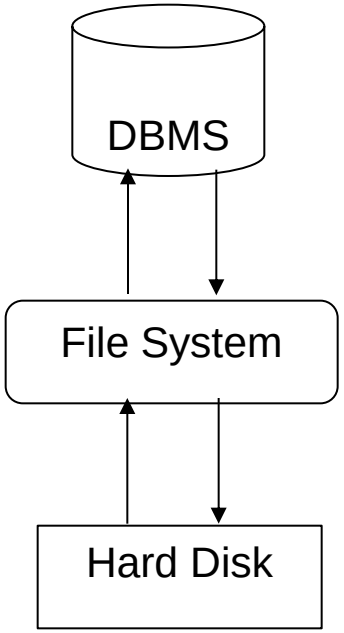


File System vs DBMS



Difference:

1. Data Redundancy
2. Data Inconsistency
3. Security
4. Atomicity Problem
5. Data Searching



Thank You