

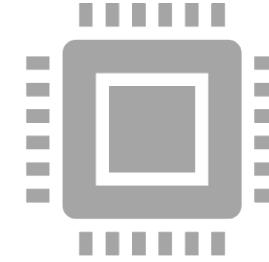
Introduction to Databases

Dr. Sara Farag



Power of
Data

Types of Databases and Database Applications



Traditional Applications:

Numeric and Textual Databases

More Recent Applications:

Multimedia Databases

Geographic Information Systems (GIS)

Biological and Genome Databases

Data Warehouses

Mobile databases

Real-time and Active Databases

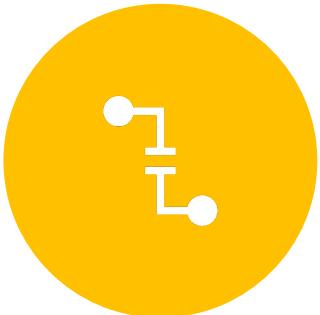
Recent Developments



New Technologies are emerging from the so-called non-database software vendors to manage vast amounts of data generated on the web:



Big Data storage systems involving large clusters of distributed computers



NOSQL (Not Only SQL) systems



A large amount of data now resides on the “cloud” which means it is in huge data centers using thousands of machines.



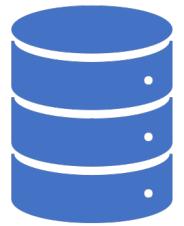
To understand what is a database we need to know the difference between data and information.



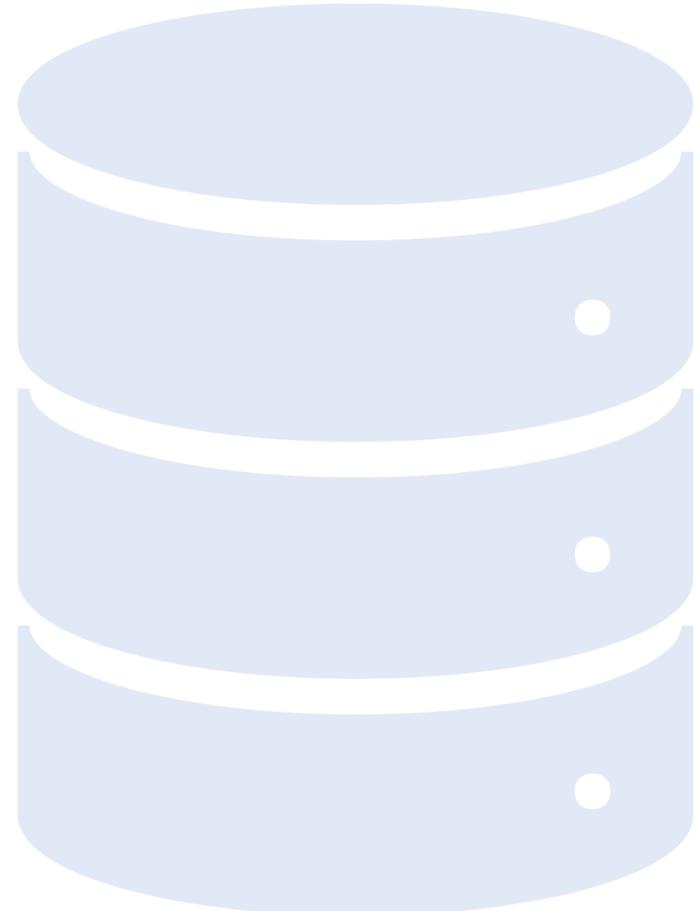
Data Vs. Information



- **Data** is raw, unorganized facts that need to be processed. Data can be something simple and seemingly random and useless until it is organized.
- Data by itself alone is not significant
- When data is processed, organized, structured or presented in a given context so as to make it useful, it is called **information**
- Information is significant by itself



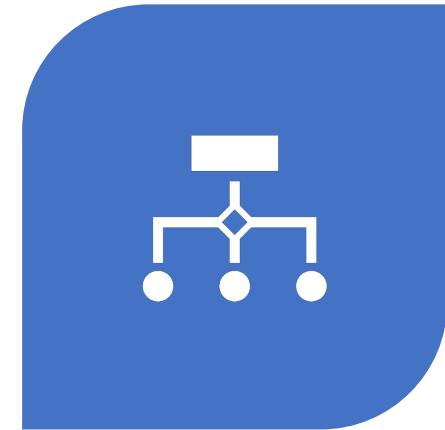
What is a database?



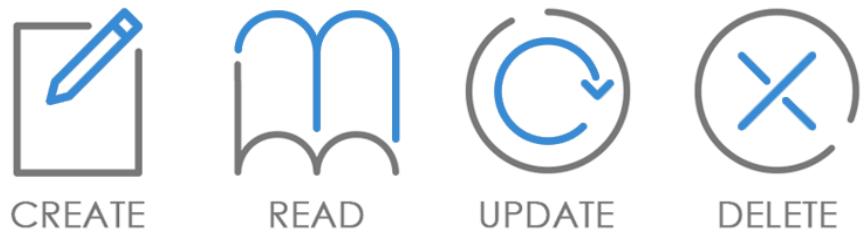
What is a database?



DATABASE (DB): A COLLECTION OF DATA THAT EXISTS OVER A PERIOD OF TIME.



THE RELATED INFORMATION WHEN PLACED IN AN ORGANIZED FORM MAKES A DATABASE.



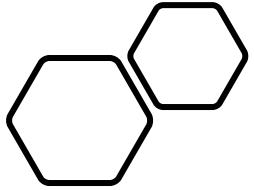
C R U D

- To add new information
- To view or retrieve the stored information
- To modify or edit the existing information
- To remove or delete the unwanted information
- Arranging the information in a desired order etc.

Operations on Databases



If you have some data
How can you store it in a computer?



Database and Computers

- There are two approaches for storing data in computers such as
 - File based approach
 - Database approach.



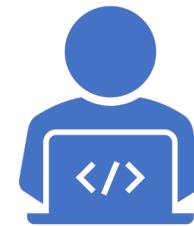
Example



Suppose we have stored in a file called Employees records having the fields/columns (emp_code, name, dept_code)



Suppose now that given an employee, for instance with name “Smith”,



we want to find out what department is he working for.

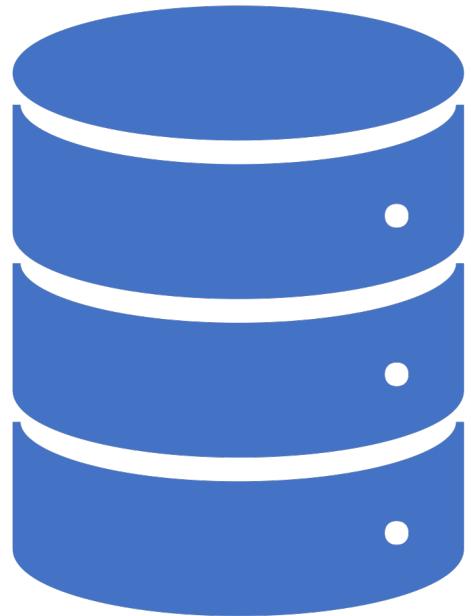
Solution

In the absence of DBMS we have to write a program which will:



1. open the file Employees
 2. declare a variable of the same type as the records stored in the file
 3. scan the file: while the end of the file is not yet encountered, assign the current record to above variable. If the value of the name field is "Smith" then remember the value of the dept_code field. Suppose it is "100"
 4. search in a similar way for a record with "100" for the dept_code in the Department file
 5. print the dept_name when successfully found the dept_code
- Very painful procedure and time consuming





Modern Solution

- Compare it to the short and elegant SQL query

```
SELECT dept_name FROM Department, Employees,  
WHERE Employees.name="Smith"  
AND  
Employees.dept_code = Department.dept_code
```



A Motivating, Running Example

- Consider building a course management system (**CMS**):
 - Students
 - Courses
 - Professors
 - Who takes what
 - Who teaches what
- Suppose that our CMS application serves 1000's of users or more **What are some Challenges?**

Relationships



Challenges with Many Users

Suppose that our CMS application serves 1000's of users or more- what are some **challenges**?

- Security: Different users, different roles

We won't look at too much in this course, but is extremely important

- Performance: Need to provide concurrent access

Disk/SSD access is slow, DBMS hide the latency by doing more CPU work concurrently

- Consistency: Concurrency can lead to update problems

DBMS allows user to write programs as if they were the **only** user

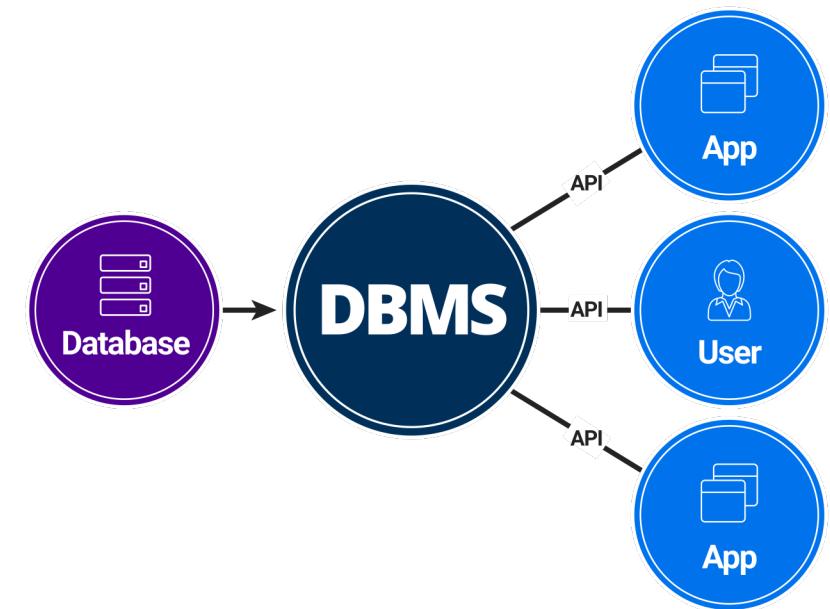
Why Not Files



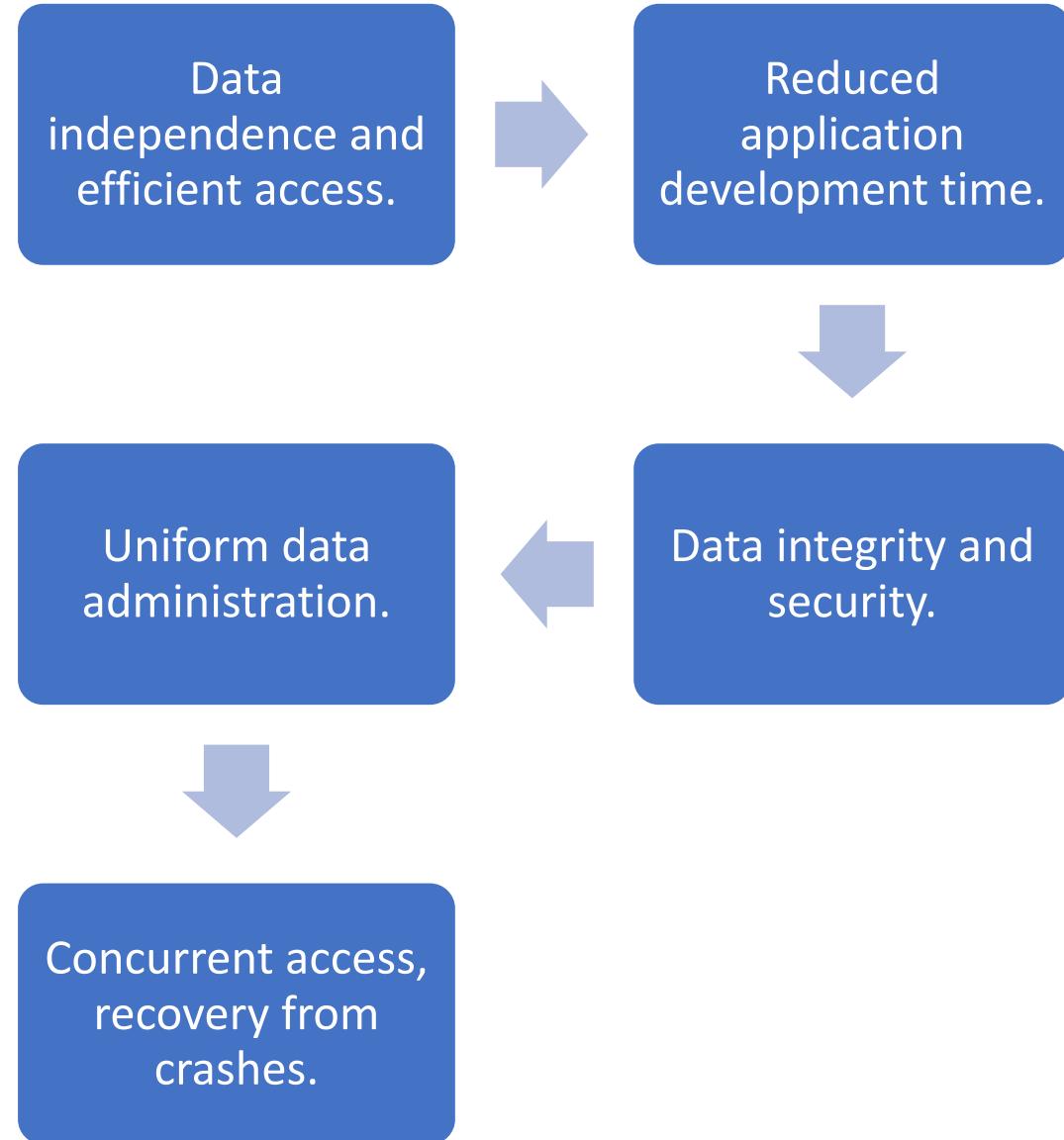
- File systems allow storage of big amounts of data
- They do not guarantee data safety(data can be lost if not backed up)
- They do not resolve an issue of modifying the same file concurrently
- No query language for the data in files.
- Need to write programs for extracting even the most elementary information from a set of files.

Such system exists

- Database Management System (DBMS) - complex software for storing and managing databases.
- A DBMS is a tool that helps develop and run data-intensive applications:
 - large databases
 - large data streams



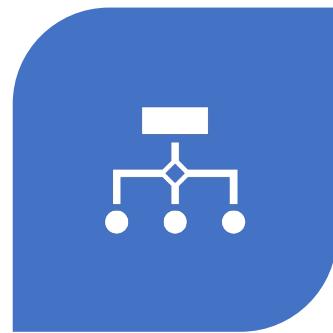
Why Use a DBMS?



What is a database?



DATABASE (DB): A
COLLECTION OF DATA THAT
EXISTS OVER A PERIOD OF
TIME.



THE RELATED INFORMATION
WHEN PLACED IN AN
ORGANIZED FORM MAKES A
DATABASE.



MANAGED BY A DBMS.