

Database Management Systems

Introduction to Databases

Malay Bhattacharyya

Assistant Professor

Machine Intelligence Unit
and
Centre for Artificial Intelligence and Machine Learning
Indian Statistical Institute, Kolkata
January, 2020

- 1 Basics
- 2 History
- 3 Data Abstraction
- 4 Languages
- 5 DBMS System Components
- 6 Limitations
- 7 Suggested Reading

Let's play a game!!!

Who can show me the student ID card first?

Let's play a game!!!

Who can show me the student ID card first?

Think why someone was ahead of others ... probably because

- it was kept at a right place (**storage**)
- it was replaced last time properly (**modification**)
- it was taken out with a fast strategy (**analysis**)

Let's play a game!!!

Who can show me the student ID card first?

Think why someone was ahead of others ... probably because

- it was kept at a right place (**storage**)
- it was replaced last time properly (**modification**)
- it was taken out with a fast strategy (**analysis**)

As a whole, we can say that the process was organized (**management**) properly by the winner.

Introduction

DBMS deals with the management of data

Introduction

DBMS deals with the management of data

Management of data refers to

- *storing* data,
- *modifying* (add, edit, delete) data, and
- *analyzing* (extract data/information) data

Note: A database is a collection of data.

Think about the past

Before DBMS, the typical file-processing systems were supported by conventional operating systems. The system stored permanent records in various files, and it needed different application programs to extract records from, and add records to, the appropriate files.

Think about the past

Before DBMS, the typical file-processing systems were supported by conventional operating systems. The system stored permanent records in various files, and it needed different application programs to extract records from, and add records to, the appropriate files.

- 1 Data redundancy and inconsistency – *repeated copies*
- 2 Difficulty in accessing data – *time complexity*
- 3 Data isolation – *changes reflected for all*
- 4 Integrity problems – *accuracy and consistency*
- 5 Atomicity problems – *everything or nothing*
- 6 Concurrent-access anomalies – *simultaneous access*
- 7 Security problems – *privacy*

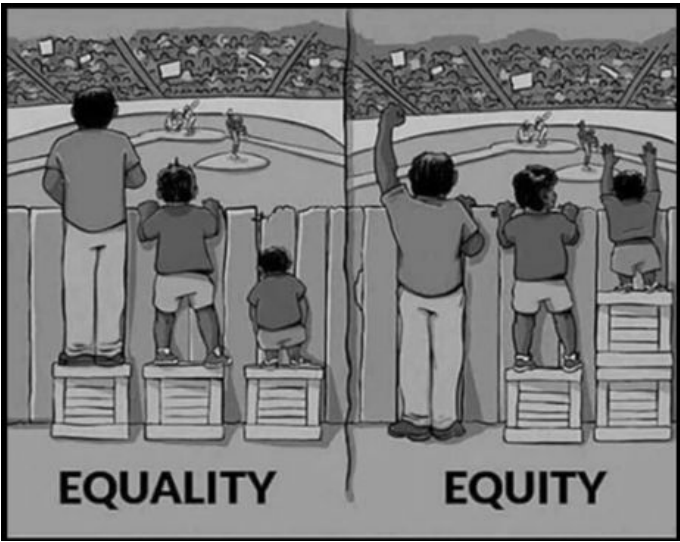
Data redundancy and inconsistency



Difficulty in accessing data



Data isolation



Integrity problems

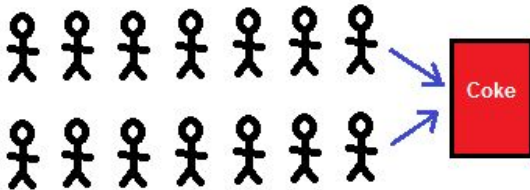


"WE'RE ALL ABOUT INTEGRITY HERE. BY THE WAY, IF MY WIFE CALLS, TELL HER I'M NOT IN."

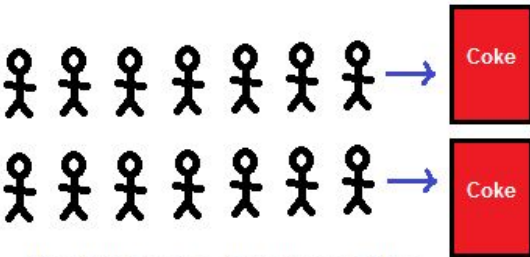
Atomicity problems



Concurrent-access anomalies



Concurrent: 2 queues, 1 vending machine



Parallel: 2 queues, 2 vending machines

History

“Data matures like wine, applications like fish” – Andy Todd.

1950s: Storage on magnetic tapes

History

“Data matures like wine, applications like fish” – Andy Todd.

- 1950s:** Storage on magnetic tapes
- Early 1960s:** Hierarchical database systems
- Late 1960s:** Network database systems
- 1970s:** Relational DBMS
- End of 1970s:** SQL
- 1980s:** Object-oriented DBMS

Let us brainstorm!!!

Suppose we wish to create a public repository to keep songs in three different raw formats – the video only, the audio, and the lyrics. The purpose is to allow the users to download these three types of files as and when required. Each of the aforementioned triplet (video, audio, text) is also associated with some metadata like the singer, year, album/movie, lyricist, etc.

Conceptualize a physical design (schema) to store the necessary data files and metadata together.

Let us brainstorm!!!

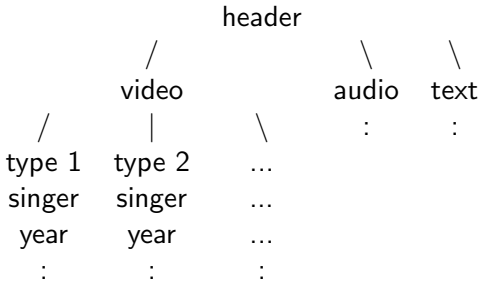
Suppose we wish to create a public repository to keep songs in three different raw formats – the video only, the audio, and the lyrics. The purpose is to allow the users to download these three types of files as and when required. Each of the aforementioned triplet (video, audio, text) is also associated with some metadata like the singer, year, album/movie, lyricist, etc.

Conceptualize a physical design (schema) to store the necessary data files and metadata together.

Note: Polyglot Persistence is a concept that encourages employing multiple data storage technologies, chosen based on the way data is being used by an application or its component, while storing data.

Idea 1

The concept: Use a hierarchical structure to organize the files and their metadata and a hierarchical structure to store the raw files.

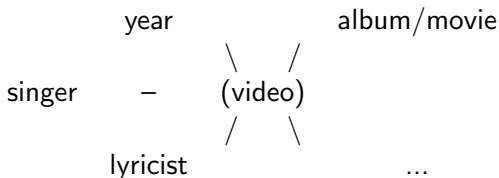


Advantages: Quick access

Disadvantages: Impractical with respect to consistency; One way searching is only possible

Idea II

The concept: Use a networked structure to organize the files and their metadata and store the raw files.



Advantages: Easy access

Disadvantages: One way searching is only possible

Idea III

The concept: Use a table to store the metadata and a hierarchical structure to store the raw files.

Song	singer	year	album/movie	lyricist	...	path
.../...

Advantages: Both way searching is possible

Disadvantages: Complex design that blends a relational and hierarchical schema

DBMS System Components

