01 (CMU)

All projects will use the CMU DB Group BusTub academic DBMS.

Architecture:
→ Disk-based storage
→ Volcano-style Query Processing
→ Pluggable APIs
→ Currently doesn't support SQL

Database

* Organized collection of inter-related data that models some aspect of the real-world

* Databases are the core component of the most computer applications

Database Examples

- GPT the year that Ice Cube went solo

Artist(name, year, country)
- "Wu-Tang", 1992, "USA"
- "Notorious BIG", 1992, "USA"
- "Ice Cube", 1989, "USA"

```
for line in file.readlines():
    record = parse(line)
    if record[0] == "Ice Cube":
        print(int(record[1]))
```

→ The code

Flat files: Data Integrity

1 - How do we ensure that the artist the same for each album entry ?

2 - What if somebody overwrites the album year with an invalid string ?

3 - What if there are muliple artists on an album ?

4 - what happen if we delete an artist that has albums ?

flate files: Implementation

1 - How do you find a particular record ?

2 - What if we now want to creat anew application that uses the same database

3 - What if two threads try to write to the same file at the same time

Flat files: Durability

1 - What if the machine crashes while our program is updating a record

2 - " " we want to replicate the database on multiple machines for high availability ?

DBMS

A database management System (DBMS) is software that allows applications to store and analyze information in a database.

A general-purpose DBMS is designed to allow the definition, creation, querying, update, and administration of database

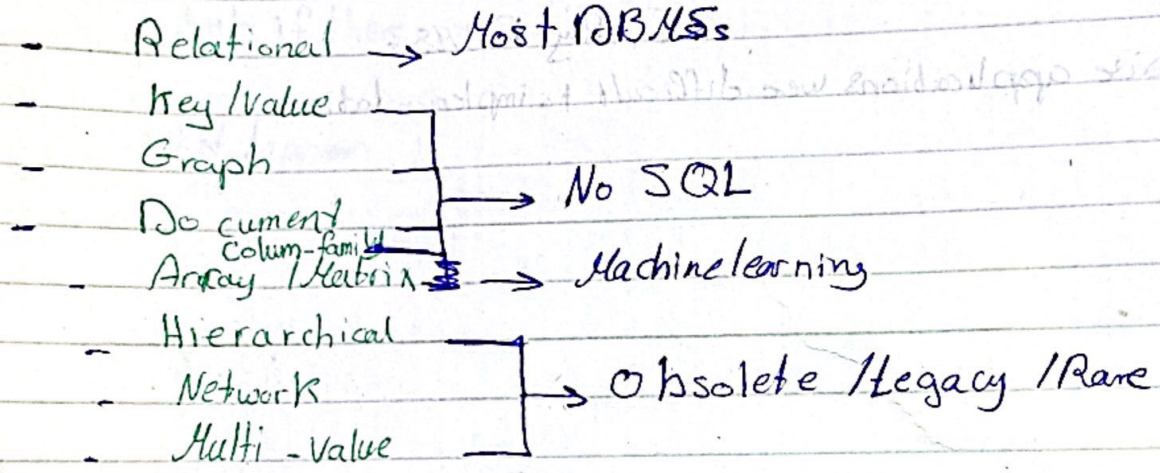
EARLY DBMSs

Database applications were difficult to implementation

Data Models

A data model is collection of concepts for describing the data in a database.

A schema is description of a particular collection of data using a given data model



Relational Model

structure: the definition of the database's relations and their contents

Integrity: Ensure the database's contents satisfy constraints

Manipulation: Programming interface for accessing and modifying a database's contents

A relation is unordered set

A tuple → the special value Null is a member of every domain

→ values are atomic

A relation's primary key uniquely

→ A foreign key specifies that an attribute from one relation has to map to a tuple in another relation

Note:

→ DML : way to store and retrieve info from DB, Two Types 1- Procedural, 2- Non procedural

→ Types of Relational Algebra :-

- select (σ)
- Projection (π)
- Union (∪)
- Intersection (∩)
- Difference (−)
- Product (×)
- Join (⋈)

Queries: The relational model is independent of any query language implementation

SQL is the defacto standard (many dialects)