Welcome Note

- COMP333: Database Management Systems
- Instructor: Bassem Sayrafi
- Office: Masri 316
- Communication
  - Email: bsayrafi@birzeit.edu
  - Facebook Group: https://www.facebook.com/groups/BZUCOMP333/
  - Office
- One of the most important courses of your careers.
- Most students can work after this course.

What is Data

- Collection of data objects and their attributes
- An attribute is a property or characteristic of an object
  - Examples: eye color of a person, temperature, etc.
  - Attribute is also known as variable, field, characteristic, or feature
- A collection of attributes describe a tuple
  - tuple is also known as record, row,entity, or an object

Attributes

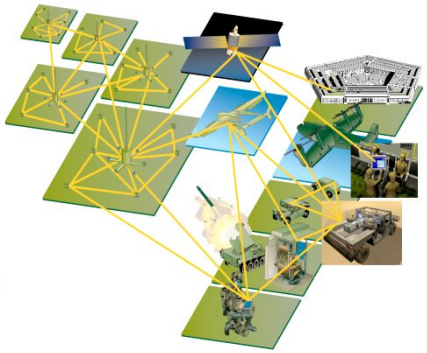
Tuples

Tid	Refund	Marital Status	Taxable Income	Cheat
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

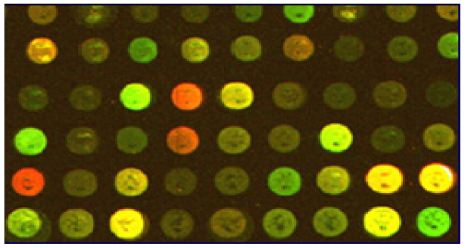
Where do we find Data?!

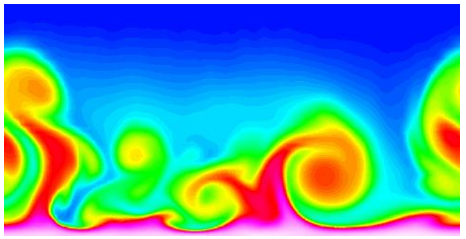
- Lots of data is being collected and warehoused
  - Web data, e-commerce
  - Financial transactions, bank/credit transactions
  - Online trading and purchasing
  - Social Network
  - Government
  - Scientific data
  - Historical
  - Measurements

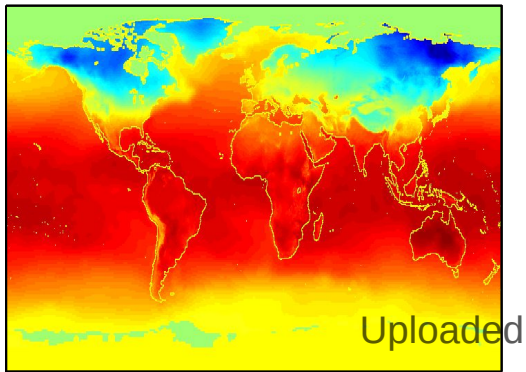
How much Data?

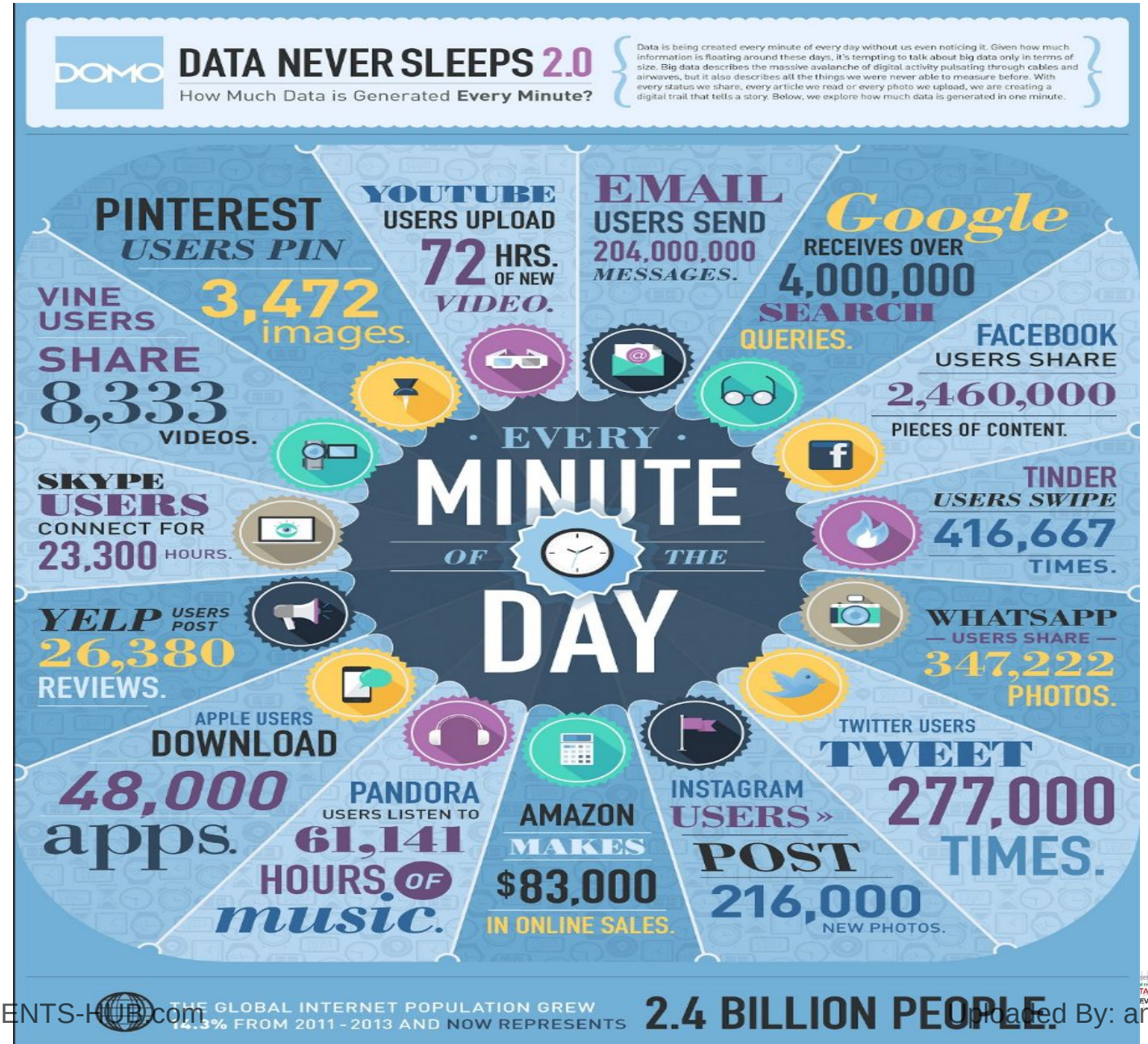


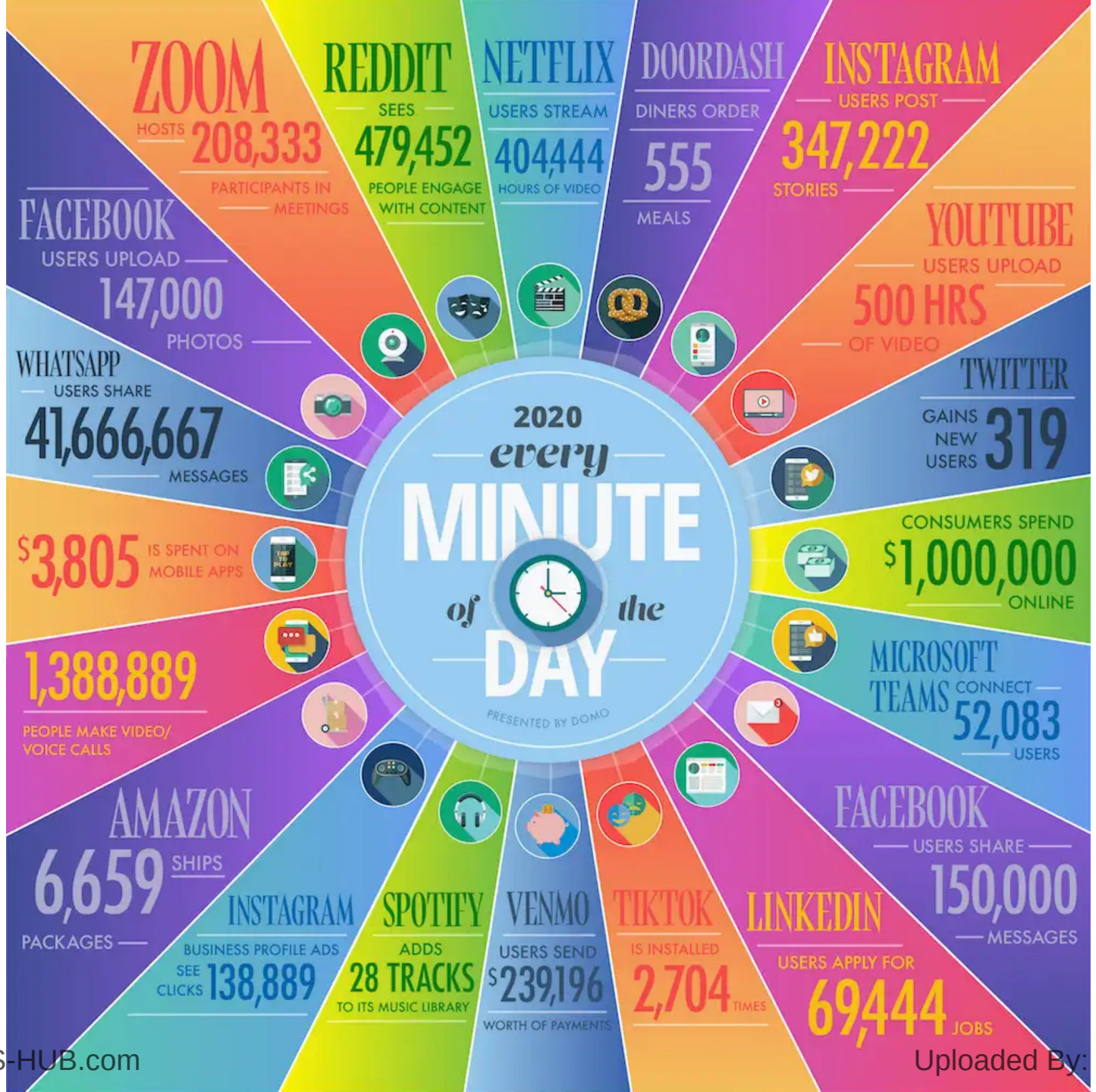
- Data collected and stored enormous speeds (GB/ho
  - remote sensors on a satellite
  - telescopes scanning the skies
  - microarrays generating gene expression data
  - scientific simulations generating terabytes of data
- Traditional techniques infeasible for raw data











How Much Social Data?

- 2016: 144.8 billion Email messages sent a day.
- 2020: 306 billion email messages sent a day

- People and brands on Twitter send more than 340 million tweets a day (revised to 500m in 2020).

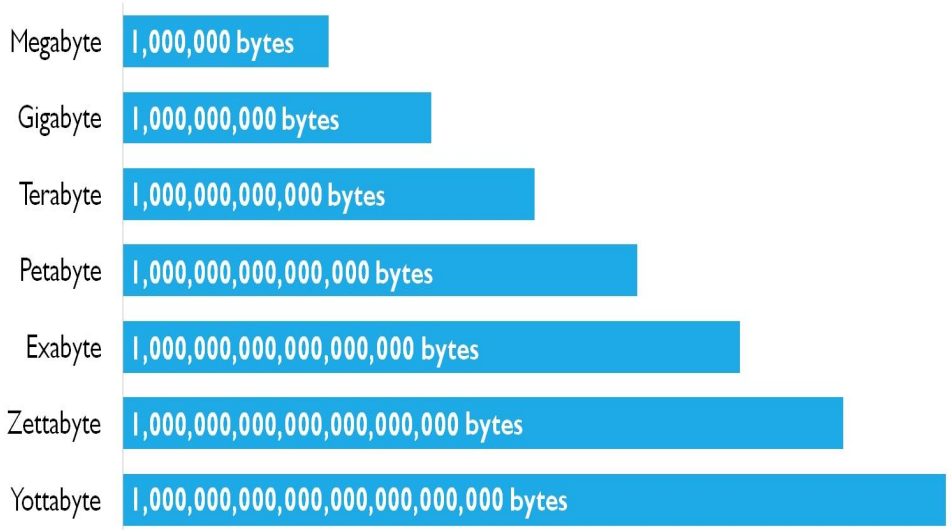
- People upload 500 hours of new video to YouTube a minute.

- Netflix: 400,000 Hour of video content daily!

  - Around 5 Billion users.
  - Average time daily around 2.5 hours.

YouTube

- 1080p (1K): Storage per minute: 50MB
- 500 hours per minute
- Total storage of 500 hour per minute: 1500GB = 1.5TB per min!
- Total storage per day: 2160 TB
- Total storage per year: 770 PT

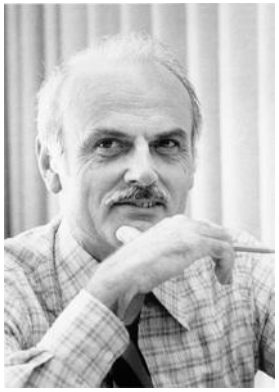


1 PB = 1000 TB
1 EB = 1000 PT
1 ZB = 1000 EB
Or
1 ZB = 1000 million TB

Crunch the Numbers

- **2008 3.6 ZB**
- **2020 44 ZB**
- **2025 181 ZB (estimate)**
- 90% of the world's data created in the last two years alone
- What does that mean?!
  - If each Terabyte in a Zettabyte were a kilometer, it would be equivalent to 1,300 round trips to the moon and back (768,800 kilometers).
- 1 gram of DNA: 215PB!!!

Back to the basics (1970)



- Need to record banking transactions
  - Withdrawal
  - Deposit
  - Transactions
  - Fail-Safe
- Edgar Codd at IBM 1970 Seminal paper
  - Proposed the relational model
  - Main idea was to organize data as groups of relations
  - Each relation describes a group of objects with similar attributes
  - ACM Turing Award

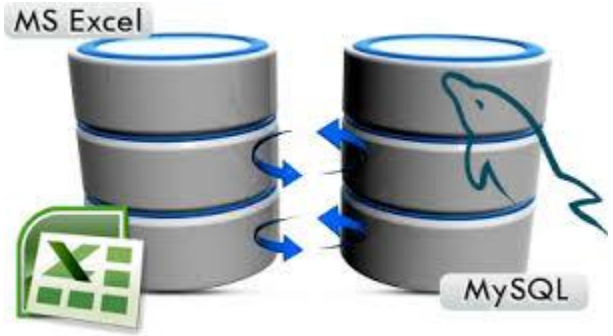
Student ID	Name	Major
1161234	Ahmad	ENCS
1161455	Noor	COMP

Course ID	CODE	Name
56478	COMP333	Database management Systems
56479	COMP232	Data Structures

DBMS

- A Database Management System (DBMS) is an integrated set of programs used to create and maintain a database.
- They are very complex systems
- Examples:
  - Oracle
  - DB2
  - MySql
  - MariaDB
  - SQL SERVER
  - Informix
- Both MySQL and MariaDB created by Michael Widenius (Monty)

Files vs. DB



- Would an Excel file suffice?!
- Suppose a population of china File!
- Problems
  - Memory
  - Questions on the data
  - Concurrent Data Access
  - Consistency
  - Security
  - Privacy
  - Recovery
  - Data depending on software!

Advantages of DBMS

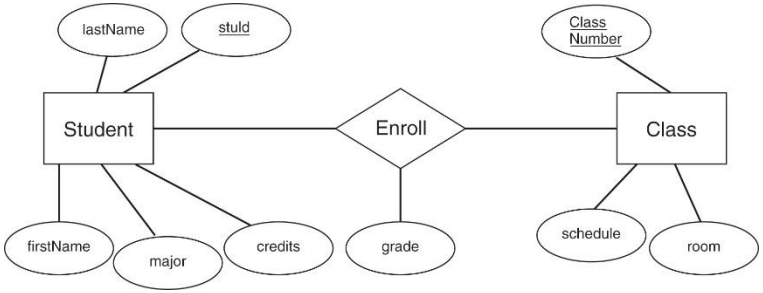
- Data independence
  - Generated and stored data should be kept separate from applications
  - The role of a database is to hold data for use by various applications. Data independence allows for the same data to be used in many different ways.
- Efficient Data Access
  - DBMS utilizes a mixture of sophisticated concepts and techniques for storing and retrieving data competently, and this feature becomes important in cases where the data is stored on external storage devices
- Data integrity and security
  - If data is accessed through the DBMS, the DBMS can enforce integrity constraints on the data.

Advantages of DBMS

- Data Administration
  - Data sharing among multiusers.
    - Integrating the administration of data can offer major improvements.
    - Experienced professionals understand the nature of the data being managed and can be responsible for organizing the data representation to reduce redundancy and make the data to retrieve efficiently.
- Concurrent Access
  - Allows for multiusers to access to the database at the same time without any problems in consistency or integrity.
- Crash Recovery
  - Protect data from system failures.
- Reduced Application Development time
  - DBMS supports many important functions that are common to many applications accessing data stored in the DBMS. This, in conjunction with the high-level interface to the data, facilitates quick development of application

Data Model: Describing Data in DBMS

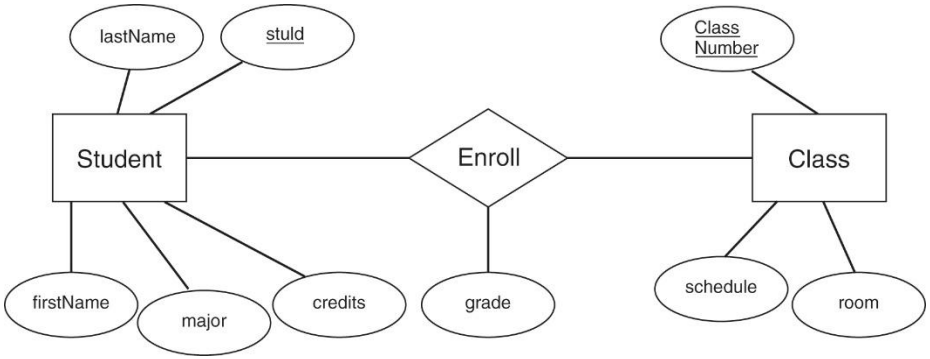
- Users ultimately concerned with model data for real-world applications.
- Data Model: collection of concepts for describing the data.
- Semantic Data Model
  - Abstract data model easier for user to describe the data of real application scenario.
  - E.g. ER Diagram
- Relational Data Model
  - Relations
  - Set of records
  - Attributes



- Schema: Description of data using a particular data model

Database Example

- A University database.
- What entities do we have?
- Entities such as
  - **Students**
  - **Teachers**
  - **Courses**
  - **Rooms**
  - **Departments**
  - **Faculties**
- Relationships between entities:
  - Students **enroll** in classes
  - Teachers **teach** courses



Relational data model example

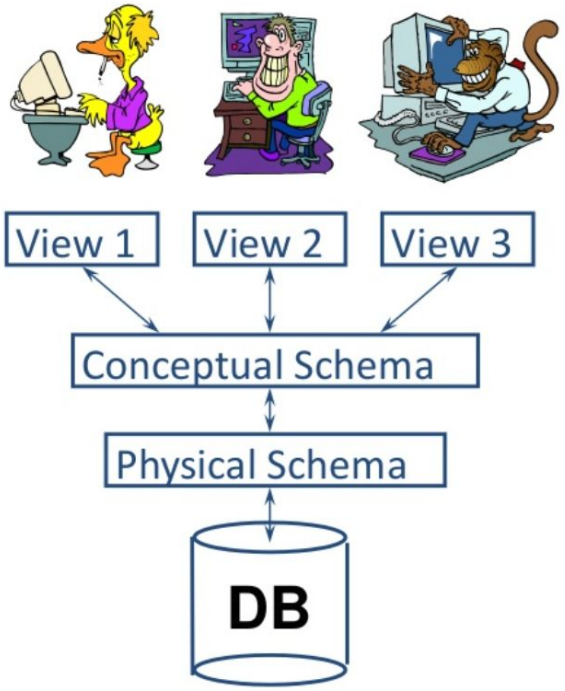
Students(*sid:* `string`, *name:* `string`, *login:* `string`, *age:* `integer`, *gpa:* `real`)

The preceding schema says that each record in the Students relation has five fields, with field names and types as indicated.[2] An example instance of the Students relation

sid	*name*	*login*	*age*	*gpa*
53666	Jones	jones@cs	18	3.4
53688	Smith	smith@ee	18	3.2
53650	Smith	smith@math	19	3.8
53831	Madayan	madayan@music	11	1.8
53832	Guldu	guldu@music	12	2.0

Levels of abstraction in a DBMS

- The data in a DBMS is described in 3 levels of abstraction
- There is a schema at each level
  - **External schema**
  - **Conceptual schema**
  - **Physical schema**



- Information about the above schemas are found in the **system catalog**

St_id	fName	lName	DOB	Major

Conceptual Level (logical level)

Student(student id int, fName: String, lName: String, DOB: Date, major: String)

Describes the stored data in terms of data model (relational)

Relational data model

- Describes all objects in terms of Entities and relationships between them.

- Quiz: Why do we store DOB rather than age?!

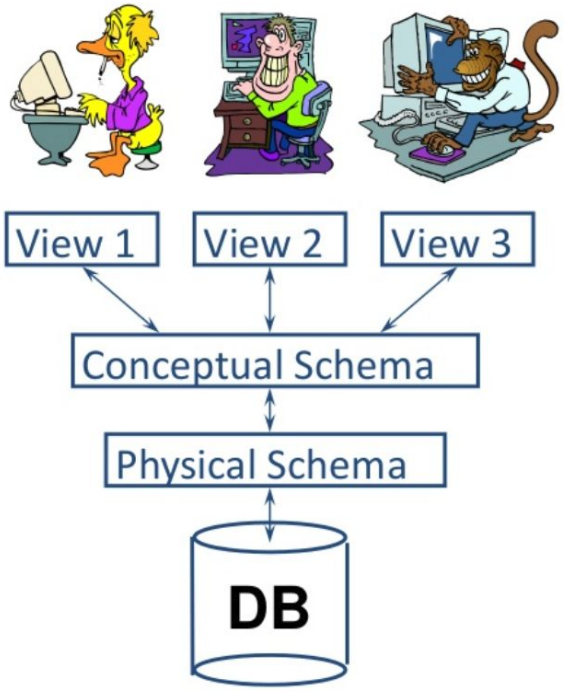
Example Conceptual Schema

Students(*sid:* `string`, *name:* `string`, *login:* `string`,
age: `integer`, *gpa:* `real`)
Faculty(*fid:* `string`, *fname:* `string`, *sal:* `real`)
Courses(*cid:* `string`, *cname:* `string`, *credits:* `integer`)
Rooms(*rno:* `integer`, *address:* `string`, *capacity:* `integer`)
Enrolled(*sid:* `string`, *cid:* `string`, *grade:* `string`)
Teaches(*fid:* `string`, *cid:* `string`)
Meets_In(*cid:* `string`, *rno:* `integer`, *time:* `string`)

Other levels

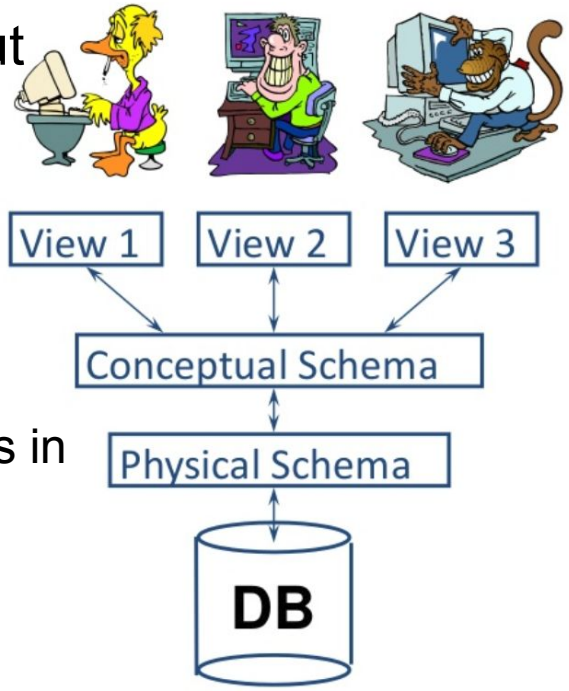
- **External level:**
  - Based on view from conceptual model
  - E.g. Ritaj

- **Physical level**
  - Concerned with how exactly the data is stored in files on storage medium



Data Independence

- Ability to modify a schema in one level without affecting the next level.
- Logical data independence
  - Users are shielded from changes in the logical structure of the data (changes in relations).
- Physical data independence
  - Conceptual schema insulates users from changes in the physical storage details.



Queries in DBMS

1. What is the name of the student with student id 123456?
2. What is the average salary of professors who teach the course with cid CS564?
3. How many students are enrolled in course CS564?
4. What fraction of students in course CS564 received a grade better than B?
5. Is any student with a GPA less than 3.0 enrolled in course CS564?

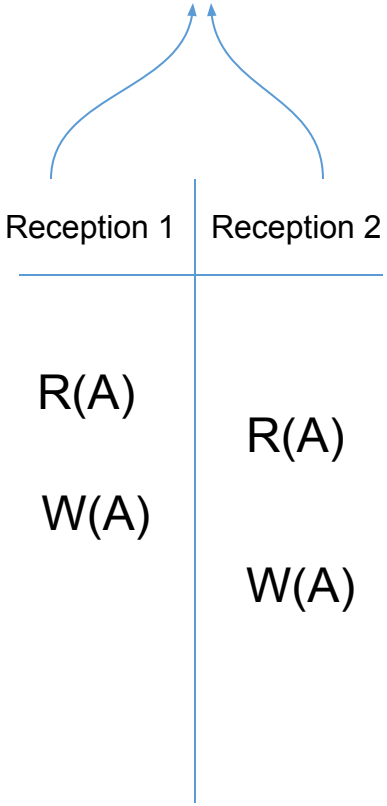
SQL

- DBMS Supports **Structured Query Language.**
  - Based on Relational Algebra

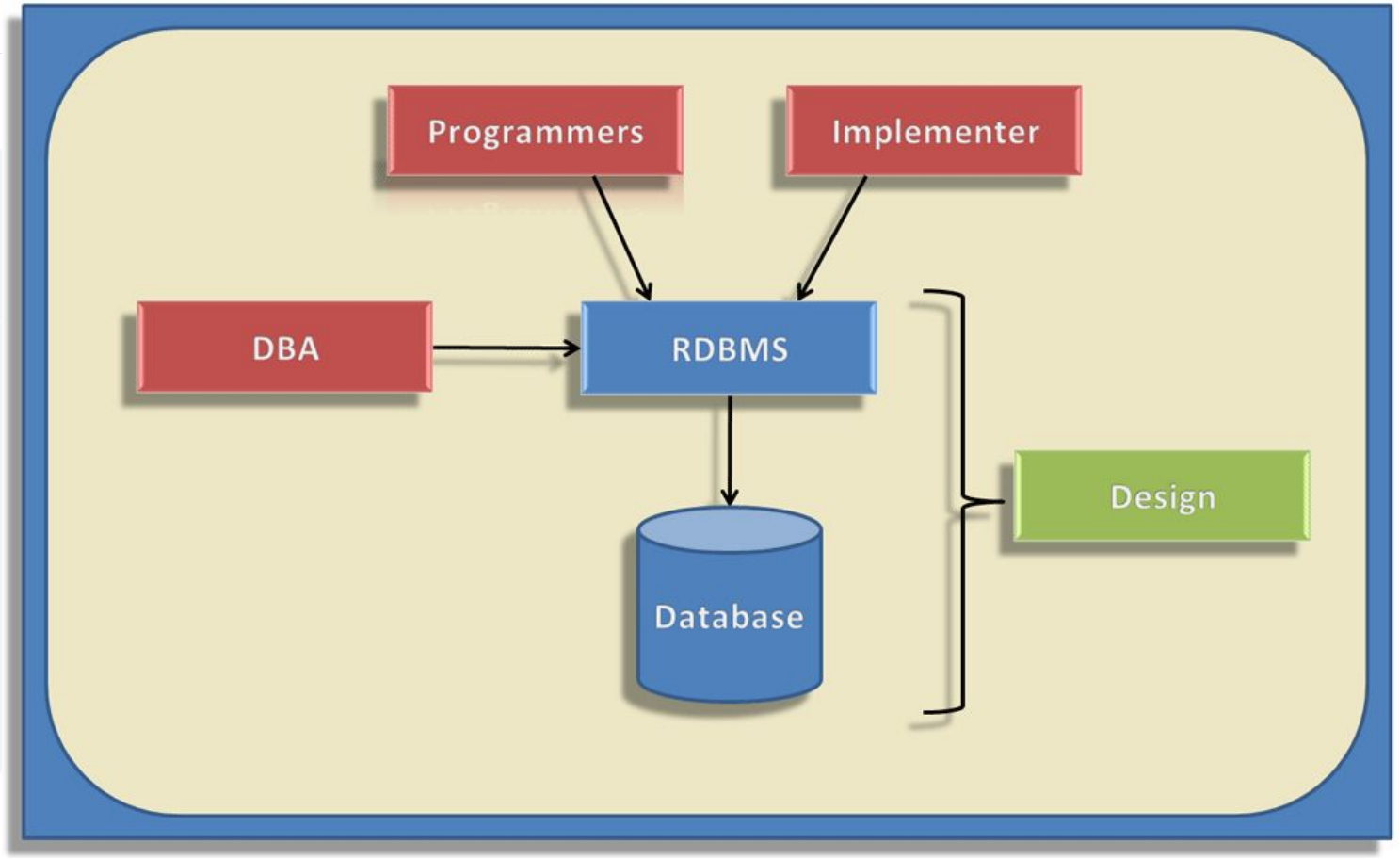
- Composed of
  - DDL
  - DML

Transaction Management

- Flight example
- Two users
  - One Ticket left!



Who works with databases?



DBMS Architecture

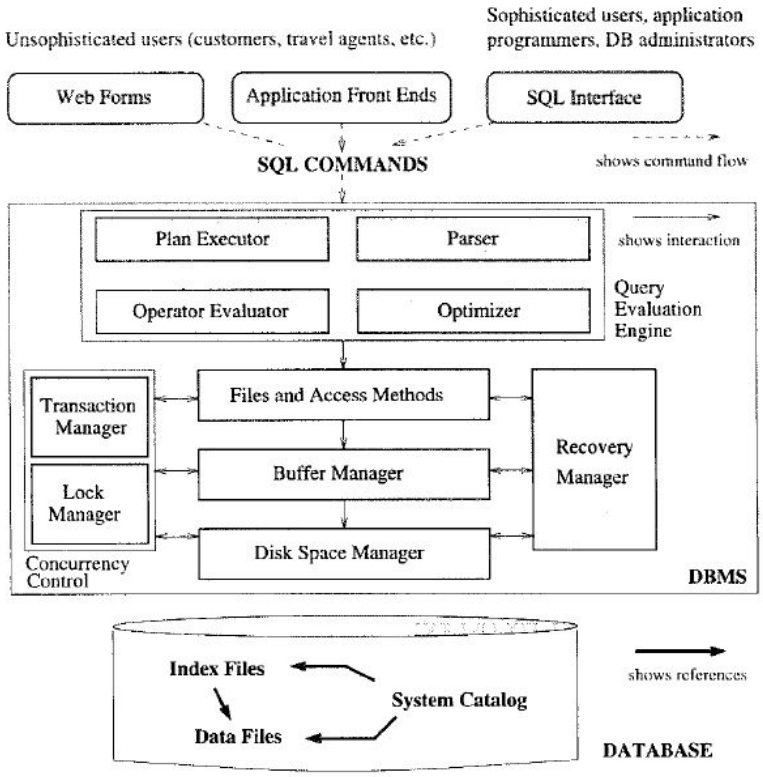


Figure 1.3 Architecture of a DBMS

What is a database?!