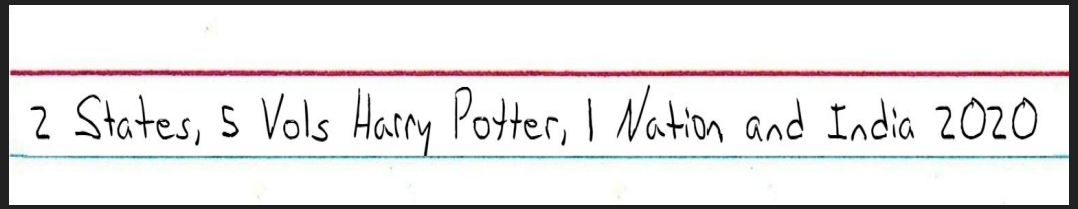

Database

Data
DataStore

- I am taking a note "what are the books I have on my bookshelf"
- Consider, I have following books
  - Two States
  - Harry Potter - 5 Vols
  - One Nation
  - India 2020
- Now I written it like below(in Paper)

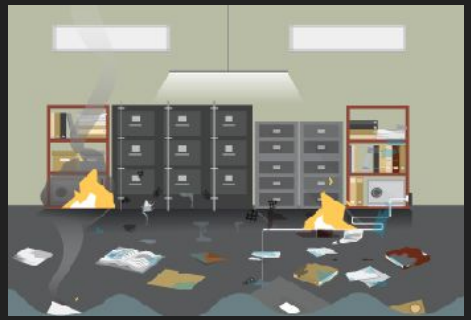


DataStore - Handwritten notes

- A paper management system can become quite the hassle.
- Also be time-consuming to organize, difficult to share with others
- Risk of being ruined or lost in the event of an unexpected disaster

Does keeping track of hundreds upon thousands of papers seem like a huge headache?

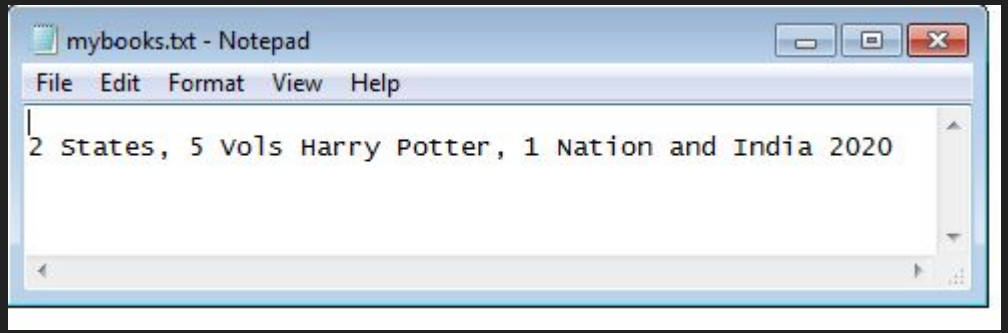
Maybe it’s time to look into an electronic document management system.



DataStore - Paper ---to--- Digital

A text file ("flatfile") is a kind of simple Digital file, that contains unformatted text (Human-readable sequence of characters and the words)

- Convenient
- Easy to share via Email and Online File sharing
- Data is protected against accidental Disaster



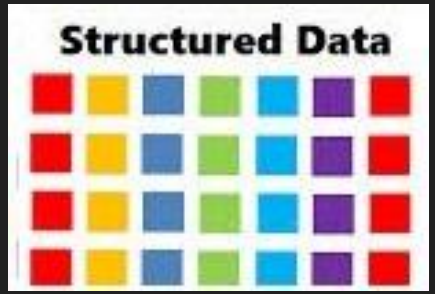
DataStore - Digital file (.txt)

“2 States, 5 Vols Harry Potter, 1 Nation and India 2020”

- In above text “2” & “1” It's not meant for quantity. But “5 Vols Harry..” meant for number of books.
- it’s not clear which word refers to what exactly.
- Unstructured has no fixed underlying structure
- might be easy to understand for a human, but for a computer this is hard to understand. This is what we call ***unstructured data***.

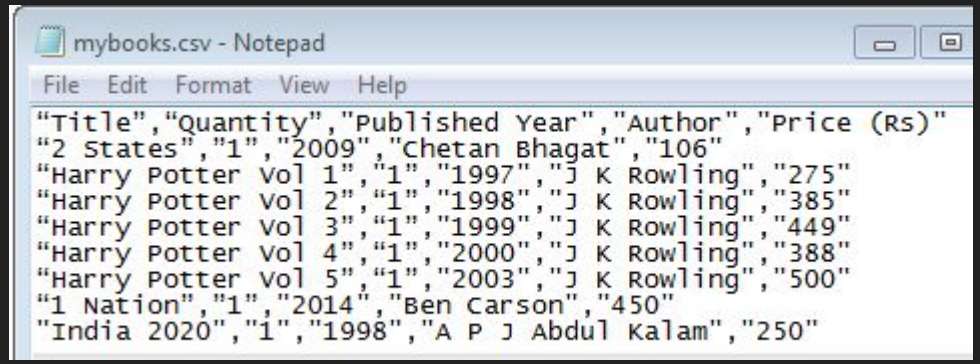
Data for Computers

- exceptionally hard to make computers extract information from certain sources
- If you want your computer to process and analyse your data, it has to be able to read and process the data.
- **It** needs to be structured and in a machine-readable form



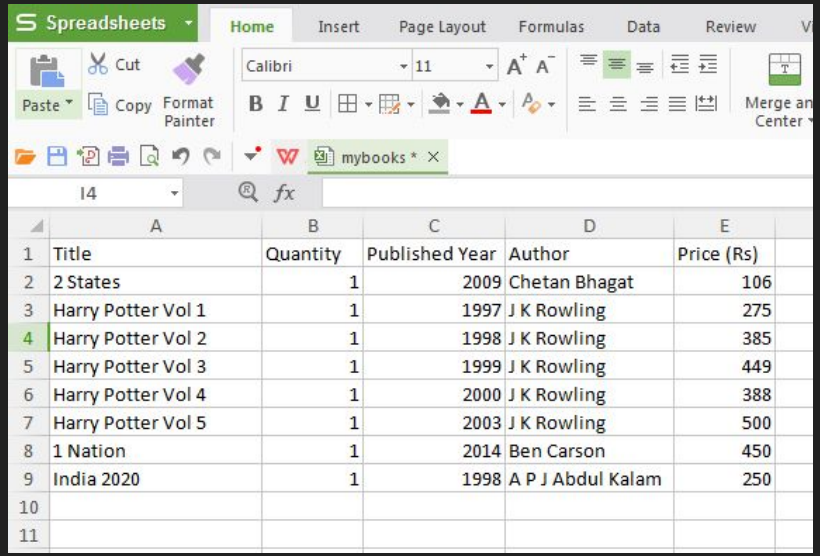
DataStore - unStructured ---to--- Structured Data

- One of the most commonly used Simple formats for exchanging data is CSV
- CSV stands for ***Comma Separated Values***
- The same thing expressed as CSV can look something like:




```
"Title","Quantity"
"2 States","1"
"Harry Potter Vol 1", "1"
"Harry Potter Vol 2", "1"
```

- Note that words have quotes around them: This distinguishes them as text (string values in computer speak)
- can be read directly by spreadsheet software.

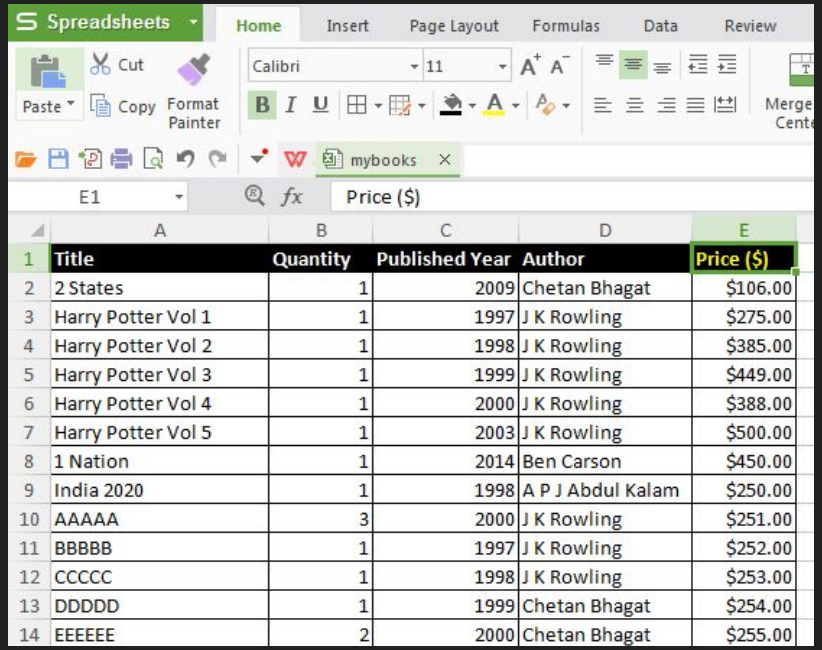


	A	B	C	D	E
1	Title	Quantity	Published Year	Author	Price (Rs)
2	2 States	1	2009	Chetan Bhagat	106
3	Harry Potter Vol 1	1	1997	J K Rowling	275
4	Harry Potter Vol 2	1	1998	J K Rowling	385
5	Harry Potter Vol 3	1	1999	J K Rowling	449
6	Harry Potter Vol 4	1	2000	J K Rowling	388
7	Harry Potter Vol 5	1	2003	J K Rowling	500
8	1 Nation	1	2014	Ben Carson	450
9	India 2020	1	1998	A P J Abdul Kalam	250
10					
11					

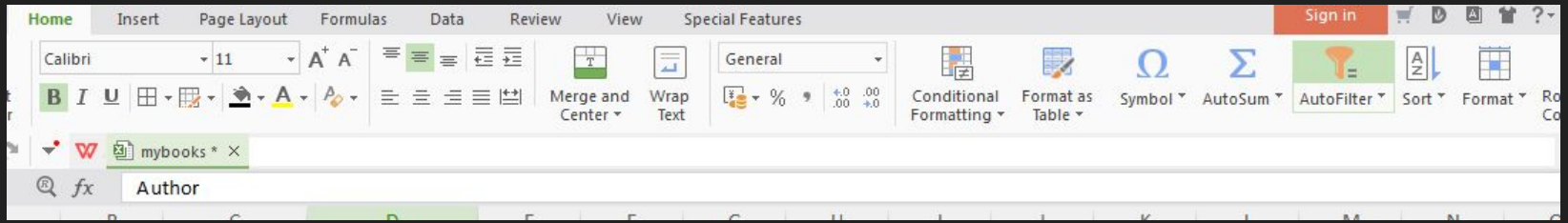
DataStore - CSV (Structured data)

Technically, This information we stored in CSV is called as "*DATA*" . CSV act as a "*DATABASE*".
It contains data that is structured in a way that's easy to Manage and retrieve.

- Another option to store the data using spreadsheet software.
- It's easy to read CSV when we have minimal data. If it's very huge, spreadsheet will be handy one
- we could do some extra things with our list(such as format it, or sort by published year, etc).
  - Sort
  - Filter
  - sum
- Program like Excel makes these tasks relatively easy to do.



	A	B	C	D	E
1	Title	Quantity	Published Year	Author	Price ($)
2	2 States	1	2009	Chetan Bhagat	$106.00
3	Harry Potter Vol 1	1	1997	J K Rowling	$275.00
4	Harry Potter Vol 2	1	1998	J K Rowling	$385.00
5	Harry Potter Vol 3	1	1999	J K Rowling	$449.00
6	Harry Potter Vol 4	1	2000	J K Rowling	$388.00
7	Harry Potter Vol 5	1	2003	J K Rowling	$500.00
8	1 Nation	1	2014	Ben Carson	$450.00
9	India 2020	1	1998	A P J Abdul Kalam	$250.00
10	AAAAA	3	2000	J K Rowling	$251.00
11	BBBBB	1	1997	J K Rowling	$252.00
12	CCCCC	1	1998	J K Rowling	$253.00
13	DDDDD	1	1999	Chetan Bhagat	$254.00
14	EEEEEE	2	2000	Chetan Bhagat	$255.00



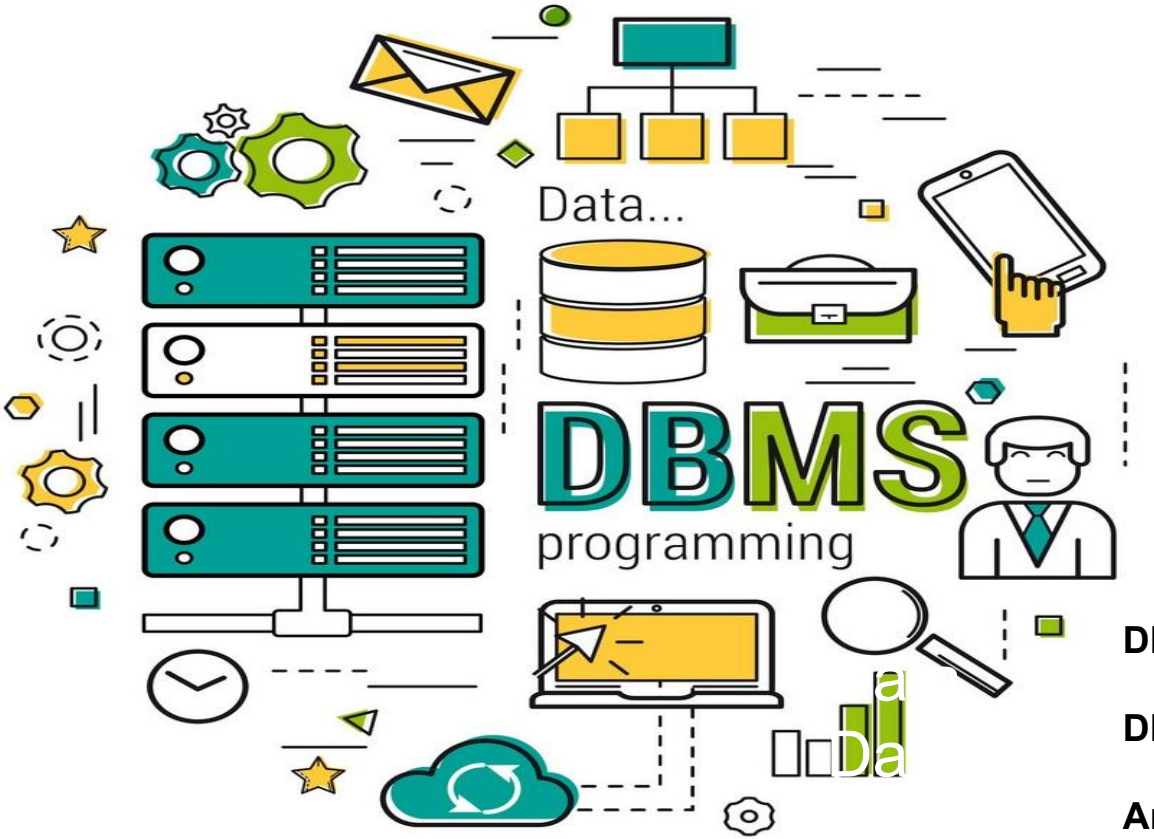
DataStore - Spreadsheet

- Specialized database software better option for Very large Data .such as MySQL, MSSQL

- Database management systems like These are purpose built for data storage and retrieval.

- The rest of this tutorial will show you why database software is a much better option for creating databases.

DataStore - Database Software



Database Management System

DBMS

DBMS History

Anatomy DB

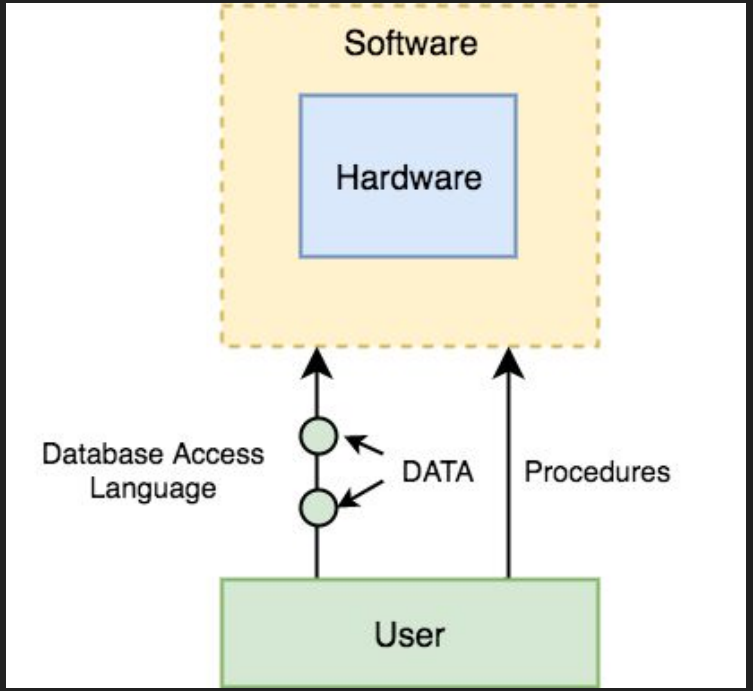
Querying

What is DBMS?

- Create, Define, Manipulate The Database
- Allow us to Store, Process, analyse data easily.
- Need to Provide Interface/Tool, data consistency & Security

Components of DBMS

- Hardware
- Software
- Data
- Procedures
- Database Access Language



DBMS - History

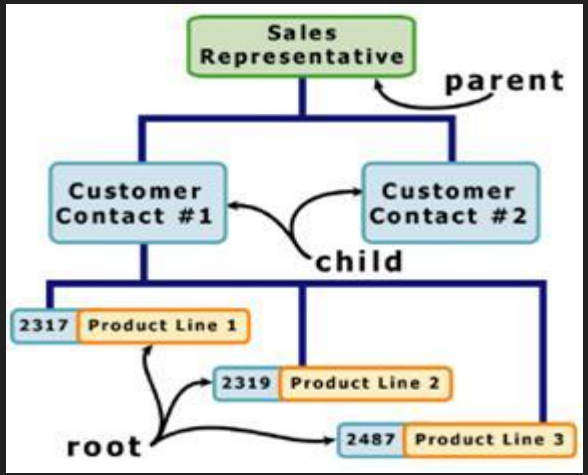
DBMSs come in many shapes and sizes. There are a variety of DBMSs available.

Major types.

- Hierarchical Databases
- Relational databases
- Object-oriented databases

Hierarchical Databases

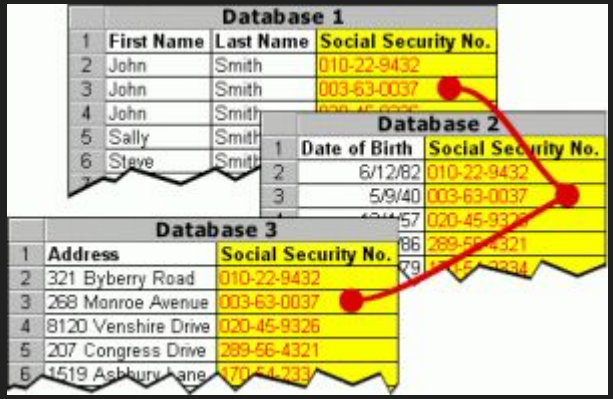
- A hierarchical database is organized in pyramid fashion, like the branches of a tree extending downwards



DBMS - Types

RDBMS

- The relationship between data files is relational, not hierarchical.
- Hierarchical and network databases require the user to pass down through a hierarchy in order to access needed data.
- Relational databases connect data in different files by using common data elements or a key field.



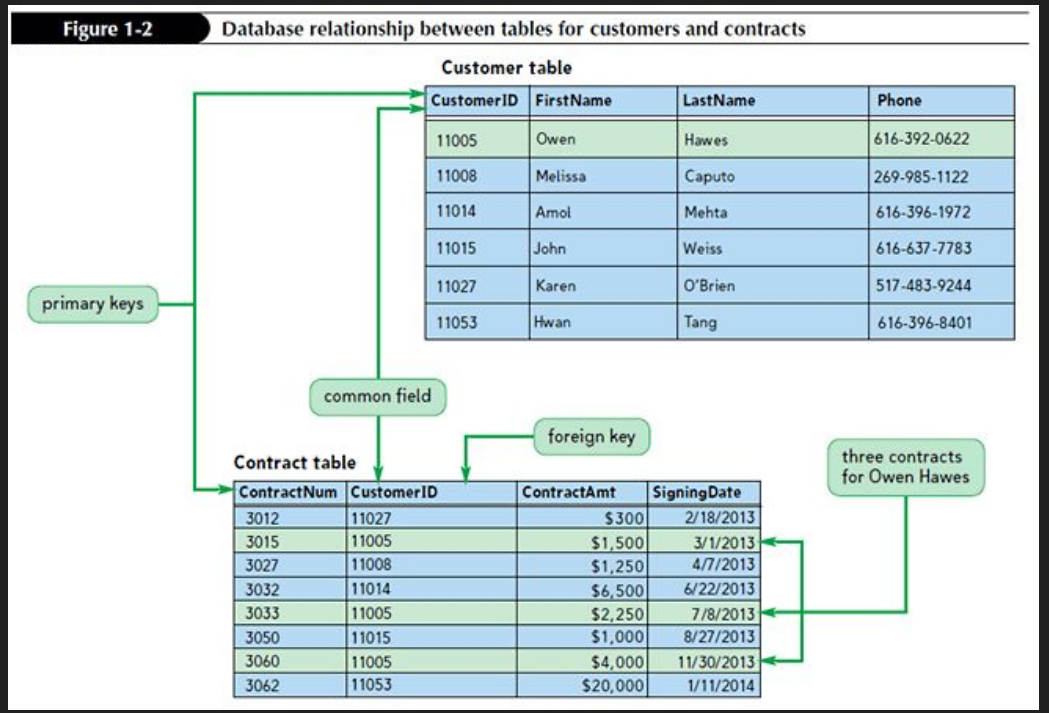
OODBMS

- New data types, including graphics, photographs, audio, and video, represent a significant advance over their other database
- All Other DBs designed to handle structured data; that is, rows and columns.
- 1) a piece of data (e.g., sound, video, text), and 2) the instructions, or software programs called methods, for what to do with the data.

DBMS - RDBMS

Collection of data organized in a structure called **table**.

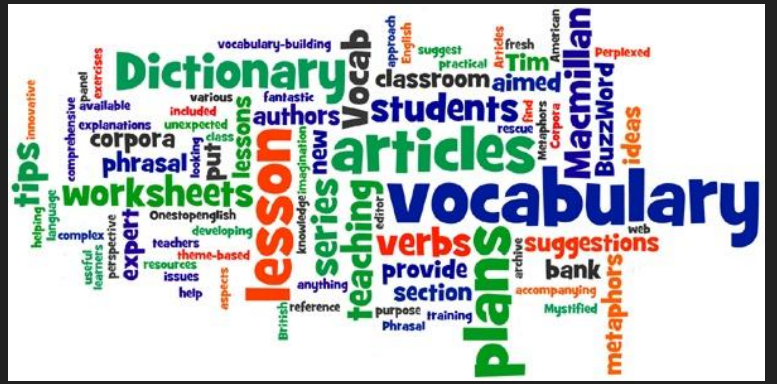
- Each table represents an entity,
- Table is a structure, made of rows and columns where **row** is a ***record/tuple*** and **column** is a ***field/attribute.***
- Relational databases connect data in different files by using common data elements or a key field.



DBMS - Anatomy DB

Vocabulary associated with DBMS

- **Data integrity** - accuracy of the data
- **Data concurrency -** The ability of a database to allow multiple users to affect multiple transactions
    - As long as the user performing the change has not saved the data, only he should be able to view the data he is changing.
    - All other users querying for the same data should view the data that existed prior to the change.
- **Data redundancy** - data that appears in multiple records
- **Security -** system, program, function and data access controls, along with the associated user identification, authentication and rights management functions



DBMS - Anatomy DB

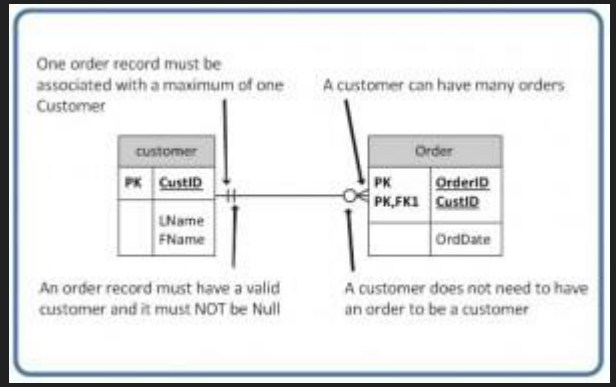
Constraints are the rules that force DBMSs to check that data satisfies the semantics. What data values are or are not allowed and which format is suitable for an attribute
Employee ID (EID) must be unique or the employee Birthdate is in the some Date range

- ***NOT NULL*** - each value in a column must not be NULL
- ***UNIQUE***- value(s) in specified column(s) must be unique for each row in a table
- ***PRIMARY KEY*** - value(s) in specified column(s) must be unique for each row in a table and not be NULL. it is used to identify individual records
- ***FOREIGN KEY*** - value(s) in specified column(s) must reference an existing record in another table (via it's primary key or some other unique constraint)
- ***CHECK***- an expression is specified, which must evaluate to true for constraint to be satisfied
- ***DEFAULT-*** Sets a default value for a column when no value is specified

DBMS - Integrity Rules and Constraints

Relation

- When creating a database, common sense dictates that we use separate tables for different types of entities.
- Example: Entity Book, Author,Publisher… etc...
- For instance, Author create Books, and Publisher contain many Books. These relationships need to be represented in the database



Several types of database relationships

- One to One Relationships
  - not very common.
  - Publisher included the address along with the Publisher could have worked fine in most cases.
- One to Many and Many to One Relationships
- Many to Many Relationships
  - we need to create an extra table
- Self Referencing Relationships

DBMS - Relationships

Determining how much redundancy exists in a table. It divides larger tables to smaller tables and links them using relationships.

- Each cell should contain a single value. record needs to be unique.
- Single column primary key
- Design to reduce the duplication of data

Querying

- Structured Query Language (SQL) is a database language designed for managing data held in a relational database management system.
- In a DBMS, the SQL database language is used to:
  - Create the database and table structures
  - Perform basic data management chores (add, delete and modify)
  - Perform complex queries to transform raw data into useful information

DBMS - Conventions