

# Intro to Database &

## Bigquery.

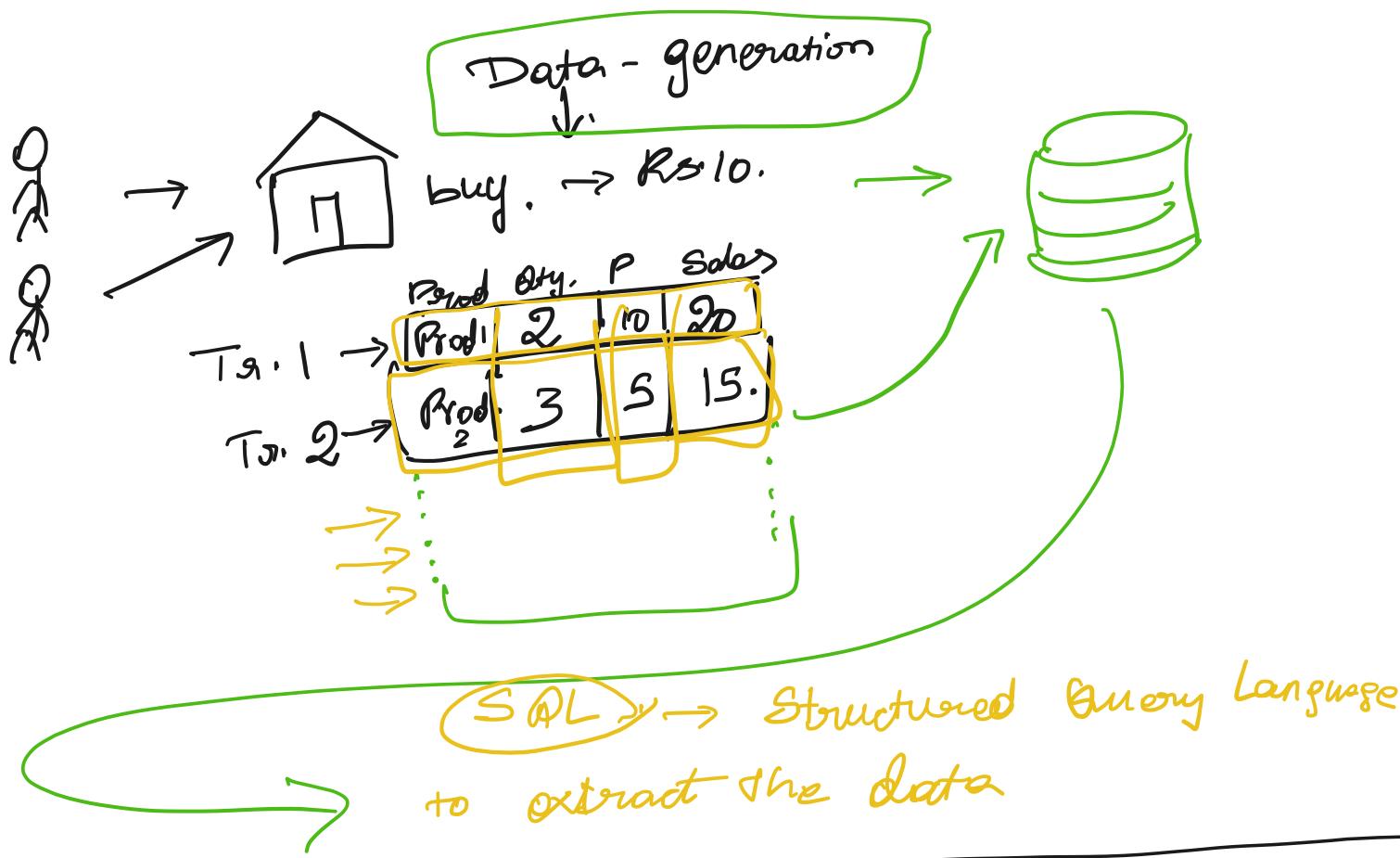
Pre-digital:

Book - Ledger

Digital

All data were stored digitally.

Where does SQL sit in Data Science space.



First thing that comes to your mind.

- ✓ Where is the data stored? DB?

- ## • What is DBMS

- What is schema
  - What are the boxes
  - What are the lines

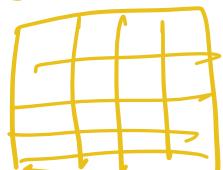
Relational DBMS called MySQL.

# Data Base Management System.

# What is DBMS?

$DB \rightarrow$  a collection of interrelated Tables.

~~MS~~ → A set of operations to manipulate the DB.  
managing



CRUD → Create, Read, Update, Delete.

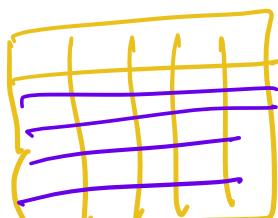
Search, Revoke, Grant, Insert, Drop.

RDBMS

- Oracle
  - MySQL
  - MS SQL
  - PostgreSQL

# structured ?

data is stored  
in row &  
column fashion  
in these DBs



Save

Save

Product table.

Pr. id	Prod. name	Price
P1	Dairy milk	5
P2	Kit Kat	10
P3	S Star	5

↓      ↓      ↑

Unstructured dBs

- MongoDB
- Cassandra
- Neo4j

NoSQL.

Unstructured data:

- Photos
- Videos
- Text (Tweets, FB post)
- Audio

Key-value format

Document format.

RDBMS

Relational DBMS

Product tab

Pr. id	Name	Price
P1	S Star	5
P2	Dairy Milk	5

Cus. table

Cus. Id	Name	Age
Cus1	John	20
Cus2	Alan	36

Sales.

Tr. id	Date	Cus. Id	Pr. Id	Qty	Sal. Amt
001	Jan 1	Cus1	P1	1	5
002	:	Cus2	P2	3	15
003	:	Cus3	P1	4	20

How much did Alan purchase? → 35

Why can't we have a single table?

Tra.id	Date	Cus.Id	Pr.Id	Qty	Sel.Amt	Name	P.Name
001	Jan 1	Cus1	P1	1	5	John	Sstar
002	:	Cus2	P2	3	15	Darryl M	
003	:	Cus2	P1	4	20	Alan	Sstar
004	:						

Assume  
 Cusid → 1KB  
 Prd id → 1KB  
 Alan → 1MB  
 Sstar → 1MB

Alan → 2MB.  
 Cus 2 → 2KB

1000 → 1000KB

- To reduce storage space
- Fast retrieval.
- Avoid redundancy

What is schema?

It is like a blue print of how the data is stored

## Data types:

The columns in tables belong to a data-type.

→ String → Name, Address.

→ Numeric → price, salary, id

↓  
int → float → decimal.

↳ whole numbers

→ date and time → to store date & time

### String datatype

Char(3) → fixed length

Varchar → variable length.

Country code <u>Char(3)</u>	Name
MAA	Thanish
BOM	John
DEL	Mickey
DXB	Hari
HYD	Jude
BLR	Michael

country code char(3) ← always 3 chars.

Name Varchar(100) ← can be max of 100 chars

### Key

Sales ↗ Hari

which Hari?

→ Okay, Hari with Cusid Ca102872

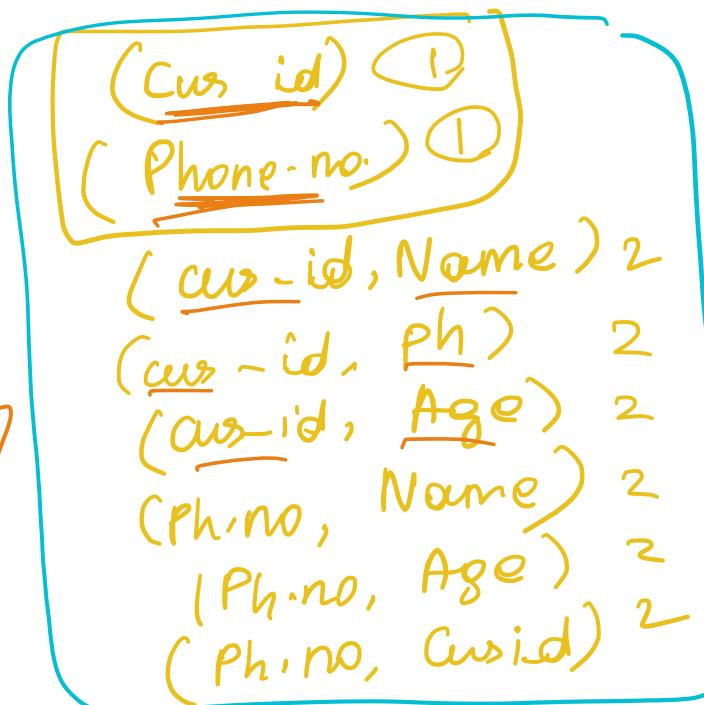
Cusid.	Name	Age	Gender	Acc.no
Cus 1	Hari			0401
Cus 2	Hari			0308
Cus 3	Alan			0920
Cus 4	John			7326,

A column that helps to uniquely identify a row is called Keys.

Examples:

Cus.id.	Name	Phone.no	Age	...
C1	John	9840	16	
C2	John	9961	17	
C3	Alan	2346	17	
C4	Mike.	9283	21	

Superkey: Single or a combination that can identify unique rows



Candidate Key: Minimal number of attributes to identify a rows i.e. from Super key.



Super key.

Candidate key  
Primary key



Primary key: • Uniquely identify a row

- Every table has 1 P.K.
- Cannot be Null.
- Cannot be duplicate



### Foreign key:

- It is a primary key in another table
- It can be Null.
- It can have duplicates

RDBMS: Relational DB.

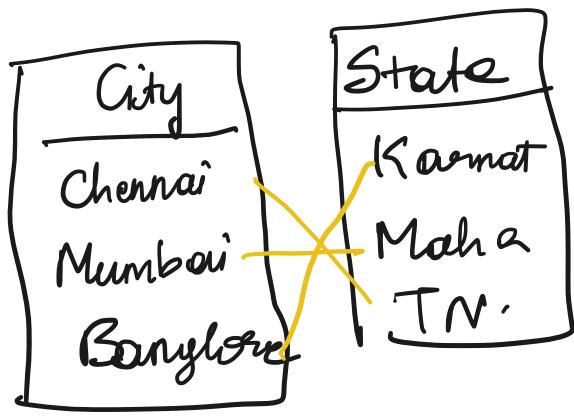
Types of relations.

• One to One

One to Many

## Many to One

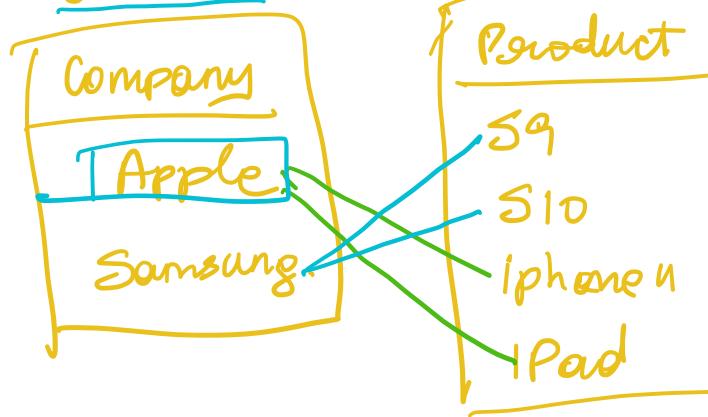
0 - 0 :



One row in one table is mapped to only one row in other tbl.

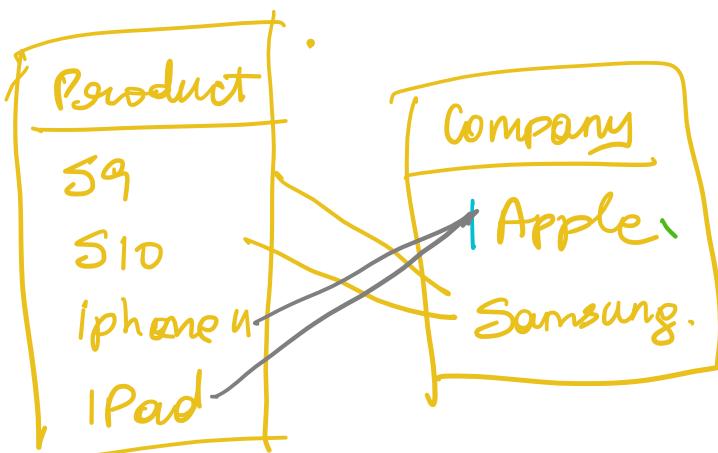
## Many to Many.

0 - to - M.

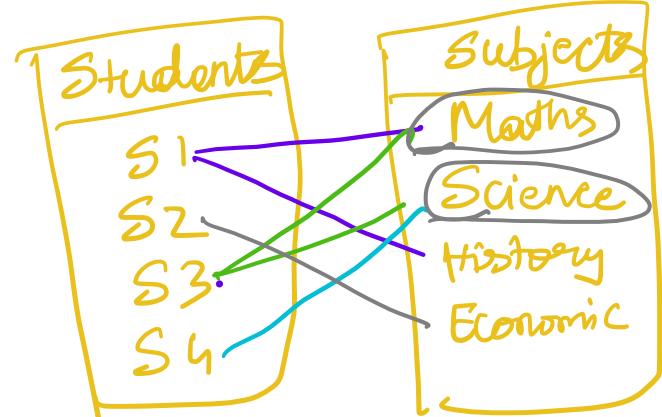


one row on lefttbl can be mapped to many rows on right table

## Many - to - one -

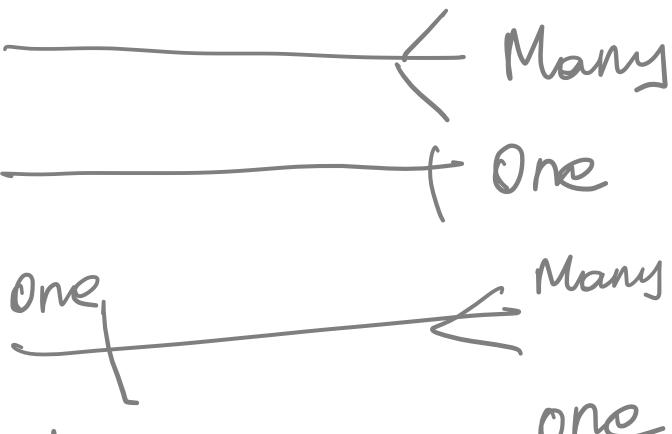


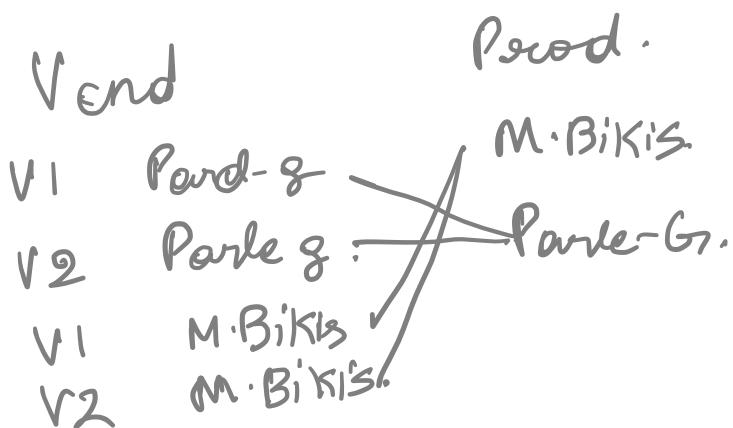
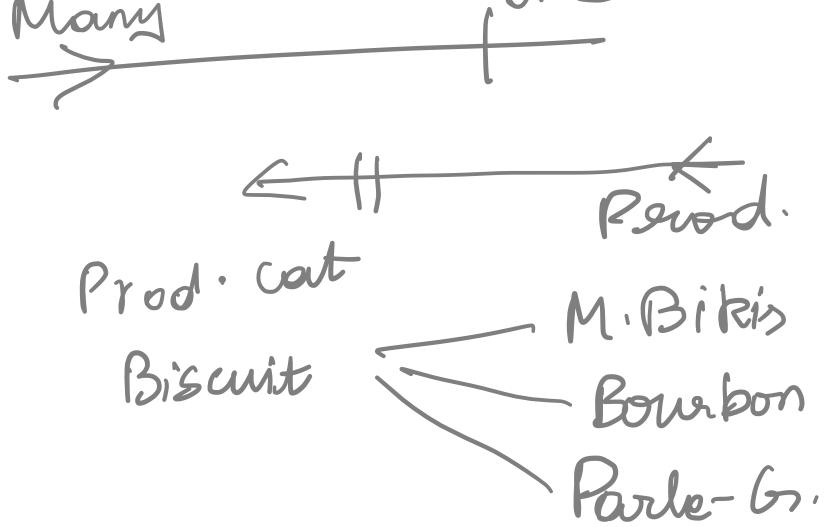
## Many to Many.



$S_1 \rightarrow \text{Math, Hist}$   
 $S_3 \rightarrow \text{Math, Science}$

$\text{Maths} \rightarrow S_1, S_3$   
 $\text{Science} \rightarrow S_3, S_4$ .





## Database

Used to collect data in realtime.

## Data Ware house

- ✓ Collection of huge Database.
- ✓ Used for analytical purpose.
- ✓ Retains historical information.
- Used by D.S / D.A. to analyse hist. data.





DW tools:

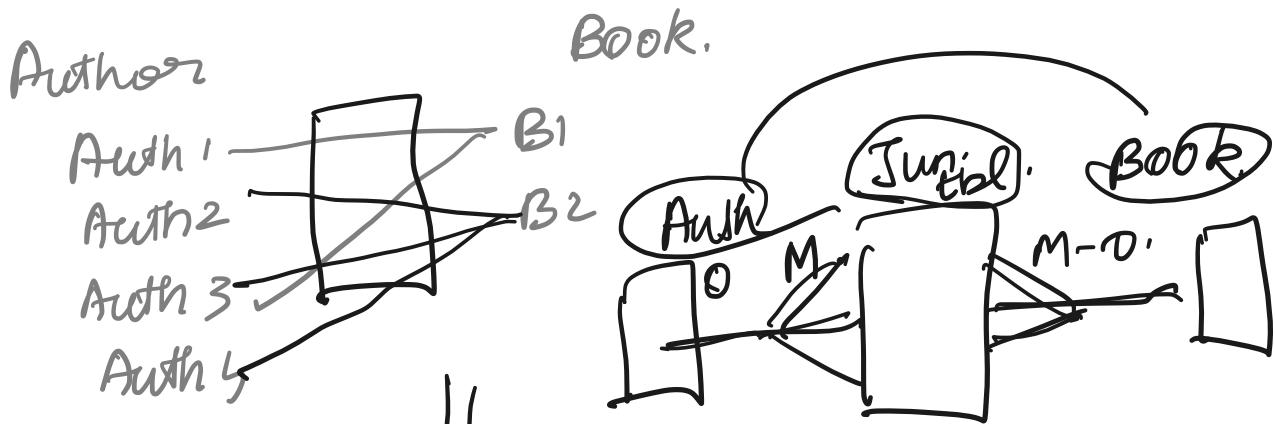
Big Query Google .

- Azure .

Redshift - AWS.

Prod id.	Date	Asstd
— AB	1	A
— AB	1	B
— AB	1	C

[ h : m : s : ms : micro : nano ]



Auth 1 B1  
Auth 2 B2  
Auth 3 B2  
Auth 4 B1  
Auth 5 B2.