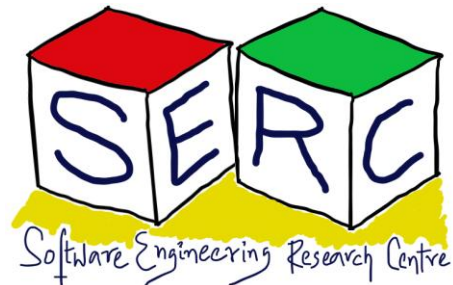


Databases

Structured Query Language(SQL)

Spring 2025

CS6.201 – Introduction to Software Systems

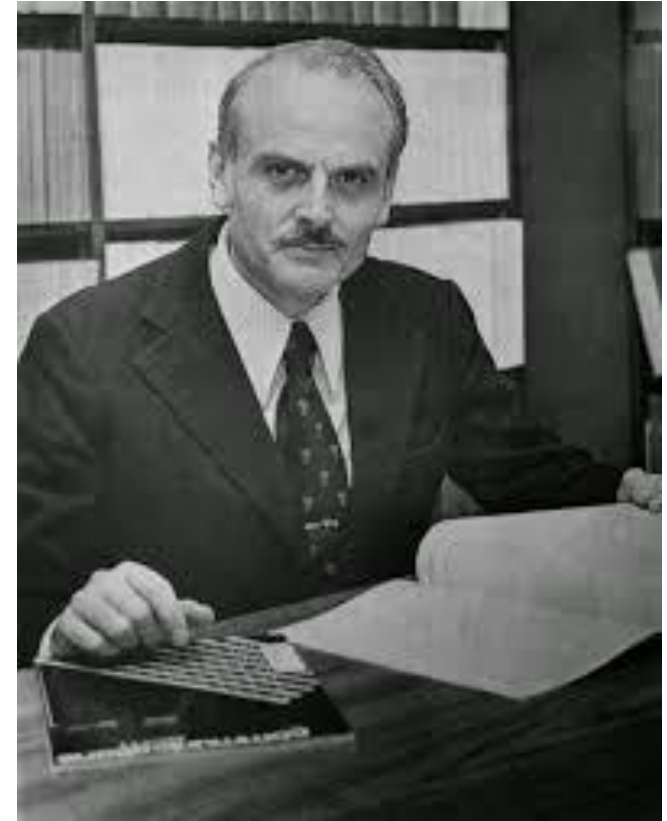




Donald D. Chamberlin



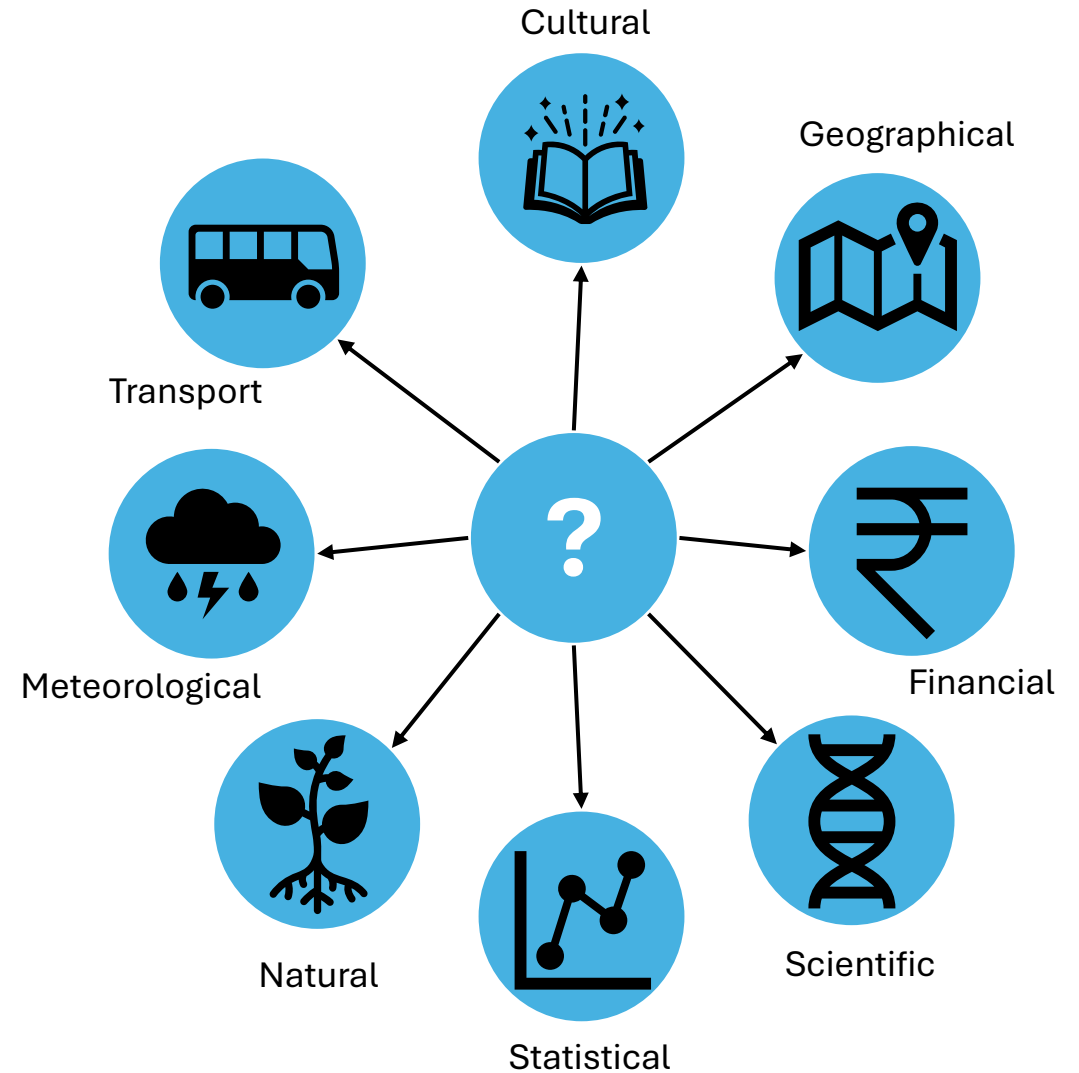
Raymond F. Boyce



Edgar F. Codd

Data

- What is data?
 - Collection of information
- Data is everywhere
- Huge amount of data is generated daily
- Personal, WWW, Organizational, Scientific
- Can we do something from this data?
 - Structured/Unstructured
 - Information and Knowledge mining



Data Explosion

- Walmart: 600+ terabytes of sales and inventory data
 - Adds a billions rows every day
 - “we know how many toothpastes are sold yesterday and what was sold along with them”
- A single astrophysics simulation of a galaxy formation can generate several PB of data, most of it thrown way
- Machine Generated data: Sensor, network devices, microphones/camera, web server logs, monitoring data etc.
- Online Services (old data)
 - X : 177M tweets sent on 3/1/2011 (nothing special about that day)
 - Dropbox: 1 M files saved every 15 minutes
 - Facebook: 135+ billion messages a month
 - Reddit: 270 million page views a month in May 2010
- Legacy Data
 - Banking & Financial transactions
 - News
 - Criminal and Motor Vehicle
 - Clinical History



<https://www.data.gov.in/>



The world's internet population continues to grow significantly year-over-year. As of November 2023, the internet represents 5.2 billion people—approximately 64.6% of the global population. According to Statista, the total amount of data predicted to be created, captured, copied, and consumed globally in 2023 is 120 zettabytes, a number projected to grow to 181 zettabytes by 2025.

As data grows and evolves, businesses need to grow and evolve, too. Domo helps you harness the power of data so you can change as quickly as the world changes and make data-driven decisions that set you apart from the crowd. Let Domo help you make sense of all the clicks, swipes, and shares so you can see the big picture that a lot of small decisions make.

Global Internet Population Growth (IN BILLIONS)



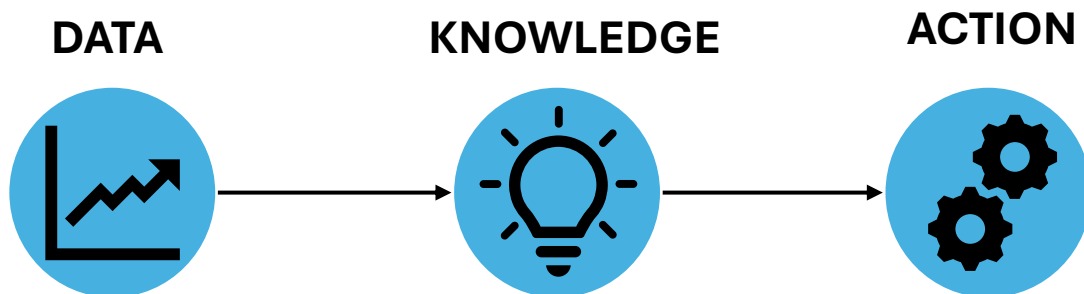
Learn more at domo.com

SOURCES: EARTHWEB, DUSTIN STOUT, DEMANDSAGE, HOOTSUITE, BUSINESSOFAPPS, DOORDASH, SOCIALPLOT, X | TWITTER.COM, GITNIX, INVGATE, THINKIMPACT, SIEMA.ORG, STATISTA, PR NEWSWIRE, NETSCOUT



More Data Explosion

- Much of this data is now stored in traditional RDMS
- **A major challenge is to manage this data, answer queries over it, gather interesting and useful insights from it.**
- **“Big Data”**
 - Everyone is either doing big data or wants to..
 - No one seems to agree on what it really means!
 - Not just about scale/volume of data
- **“Data Scientist”**
 - Goal: Extract meaning from data and creating data products
 - Need a board range of skills: programming, statistics, math....



Organized collection
(large and related) of
data.



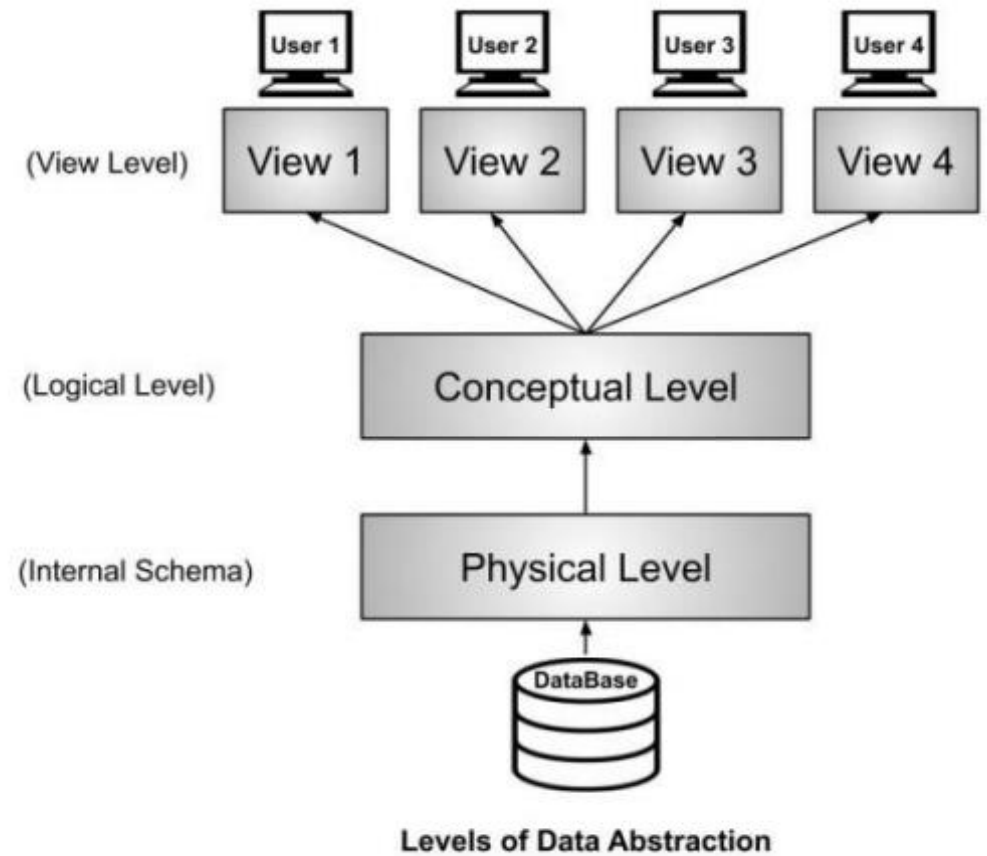
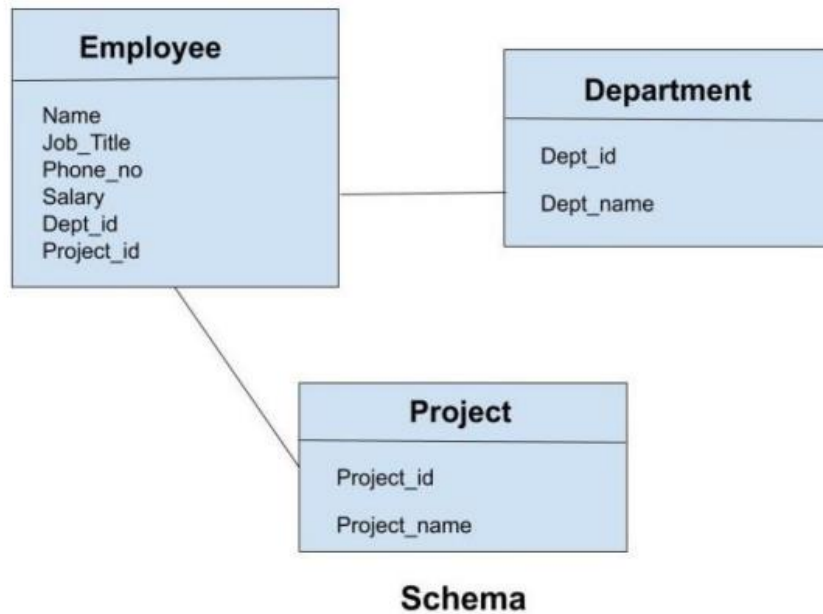
Why database?



Analyze, manage and
use data efficiently.



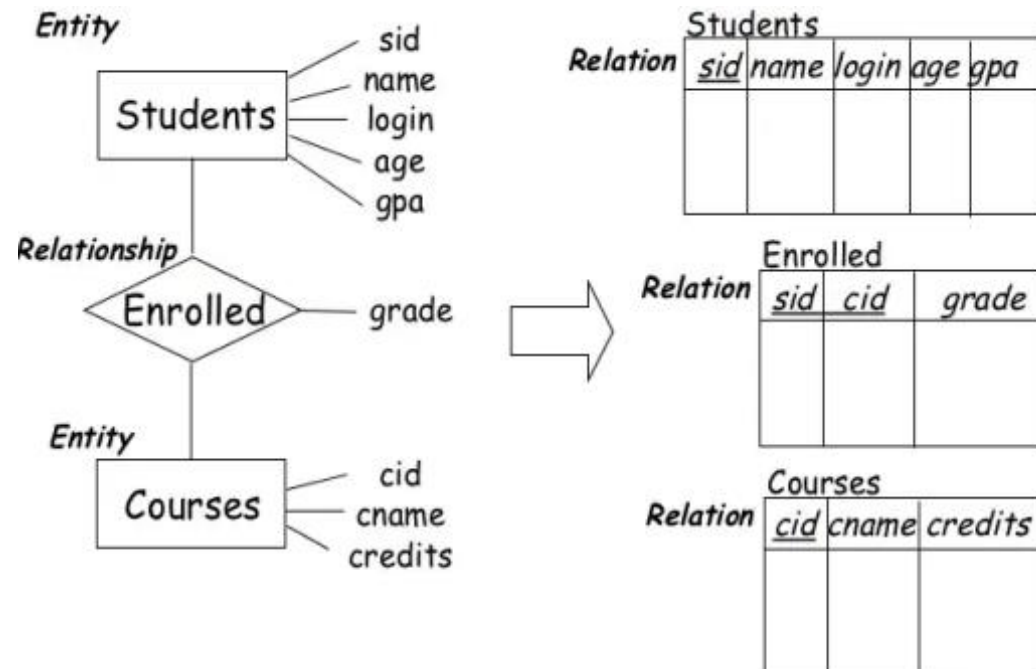
Data is accessed,
manipulated and
updated.



Physical level: describes how a record is stored

Logical level: describes data stored in database, and the relationships among the data

View level: application programs hide details of data types. Views can also hide information (such as an employee's salary) for security purposes.



Logical data independence: Protection from changes in logical structure of data.

If a user application operates on a subset of attributes of a relation, it should not affect when new attributes are added. Can make changes without affecting existing schemas.

Physical data independence: Protection from changes in physical structure of data.

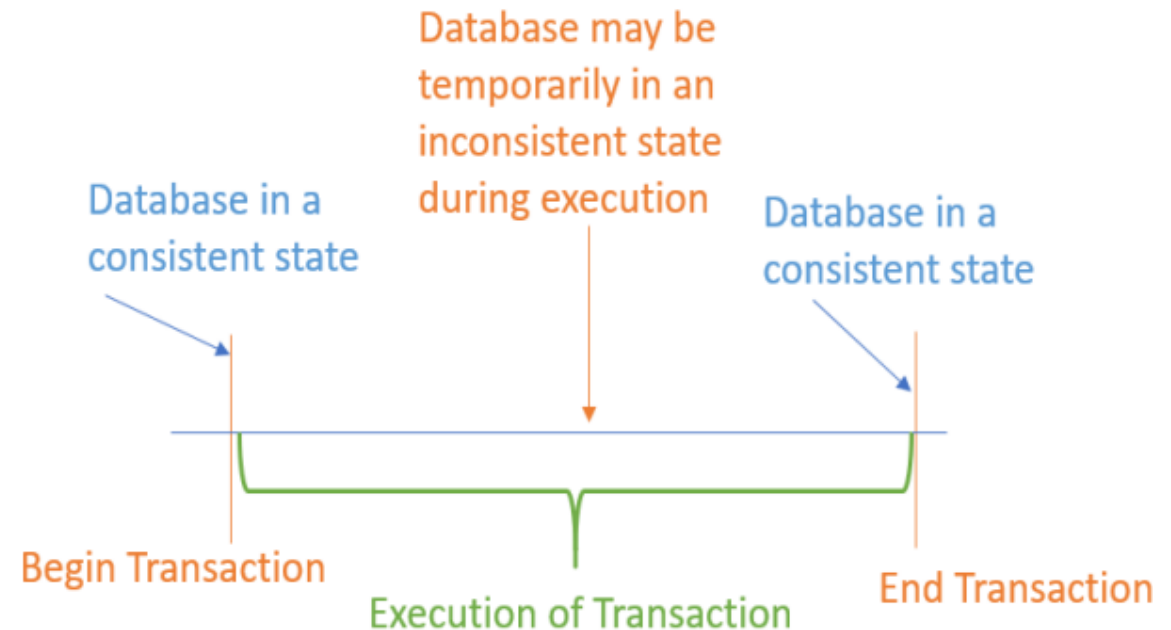
Change in internal data using different file organization or devices should be possible without changing logical schema structure.

Atomicity: If any operation is performed on the data, either it should be performed or executed completely or should not be executed at all.

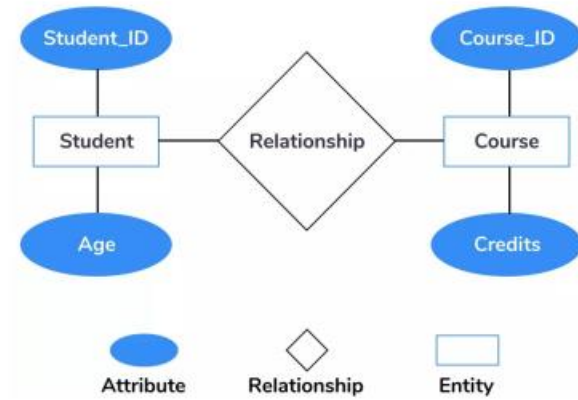
Consistency: Integrity of the data should be maintained, which means if a change in the database is made, it should remain preserved always.

Isolation: Transactions occur independently without interference. Changes should be visible only after they have been made to the main memory

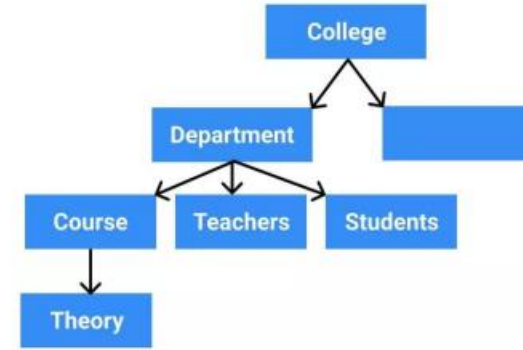
Durability: Data after the successful execution of the operation becomes permanent in the database.



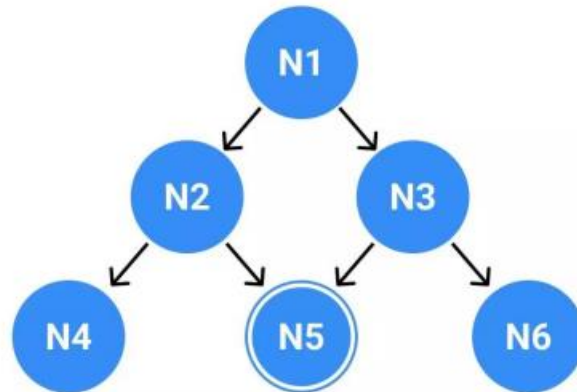
Entity Relationship Model



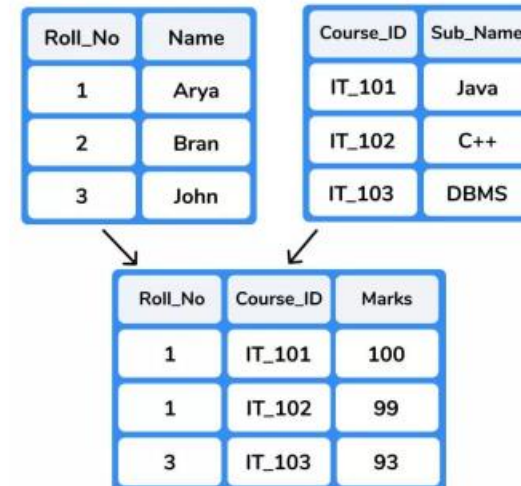
Hierarchical Model



Network Model



Relational Model



Data Point | Data Model | Dataflow Design | Data Transfer | Data Maintenance

Data Definition Language – Define and Manage

CREATE, ALTER, DROP, TRUNCATE, RENAME, COMMENT

Data Manipulation Language – Manipulates data

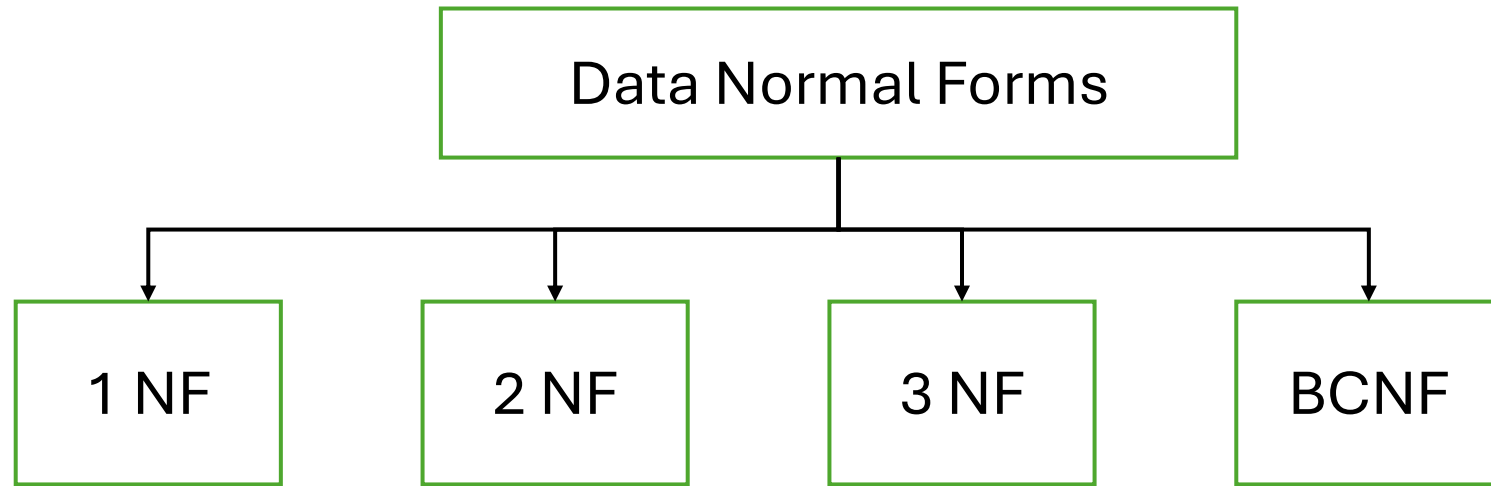
SELECT, INSERT, UPDATE, DELETE, MERGE, CALL, EXPLAIN PLAN, LOCK TABLE

Redundancy

Insertion anomalies: It may be impossible to store certain information without storing some other, unrelated information.

Deletion anomalies: It may be impossible to delete certain information without losing some other, unrelated information.

Update anomalies: If one copy of such repeated data is updated, all copies need to be updated to prevent inconsistency.



- Process of organizing the data in the database.
- Reduces data redundancy and eliminates undesirable characteristics like Insertion, Update and Deletion Anomalies.
- Normalization rules divides larger tables into smaller tables and links them using relationships.
- Normal form is used to reduce redundancy from the database table.

1NF : Each Table
cell should
contain a single
value, and each
record should be
unique

Student#	Advisor	Adv-Room	Class1	Class2	Class3
1022	Jones	412	101-07	143-01	159-02
4123	Smith	216	201-01	211-02	214-01

Student#	Advisor	Adv-Room	Class#
1022	Jones	412	101-07
1022	Jones	412	143-01
1022	Jones	412	159-02
4123	Smith	216	201-01
4123	Smith	216	211-02
4123	Smith	216	214-01

2NF: Stay in 1NF,
all non-key
attributes are
fully functionally
dependent on
the primary key

Students:

Student#	Advisor	Adv-Room
1022	Jones	412
4123	Smith	216

Registration:

Student#	Class#
1022	101-07
1022	143-01
1022	159-02
4123	201-01
4123	211-02
4123	214-01

3NF: Stay in 2NF, Has
no transitive functional
dependencies
(Changing a non-key
column, might cause
any of the other non-
key columns to change)

Students:

Student#	Advisor
1022	Jones
4123	Smith

Faculty:

Name	Room	Dept
Jones	412	42
Smith	216	42