Data and Data Flow

- Data
  Collection of information that can be stored, processed, and analysed to generate insights and knowledge. This information can be in various formats, including text, numbers, images, audio, and video.

- Data Representation at different layers in different formats -

    a. Business Layer - Text, Videos, Images etc.
    b. Application layer - JSON / XML
    c. Data Stores (Databases) - Tables, Indexes, Trees
    d. Network Layer - Packets
    e. Hardware Layer - 0s and 1s

- While designing systems we need to focus on different formats and different properties of data so that data flows efficiently.

- Examples of data stores-
    a. Databases
    b. Queues
    c. Caches
    d. Indexes

- Data flow methods -
    a. APIs
    b. Messages
    c. Events

- Data Generation -
    a. Users - Interact with systems and create more data
    b. Internal data - A type of data which systems populate on their own
    c. Insights - When a user interacts with a system and uploads / retrieves data, some insights are generated.

- Factors to Consider to help build systems :-
    a. Type of data - Knowing the type of data that we operate on is important as it would lead to help in making various decisions like which database to use to store videos vs which database to use to store some text configurations.

b. Volume of data - we need to consider the volume of data as if a system deals with terabytes of data it would be different from a system that deals with less data.

c. Consumption/retrievals - Number of reads and writes could help decide what kind of data store we need to use for our system, like some systems may have a lot of write operations but consumption of data is less and vice versa.

d. Security - Security requirements for different systems could be different like in transactional systems where security is of utmost importance.

Examples for having different types of data and data flow requirements -

- Authorization systems - Volume of data may not be high because it has to store some user details and credentials etc. but the level of security needed in an authorization system is comparatively high.
- Streaming systems - Data volume is high and data retrieval speed is high (number of requests coming to system in order to retrieve data).
- Transactional systems - The journey of a transaction is of utmost importance like it should not fail, it should not be done twice, money in the bank should be intact.
- Heavy Compute Systems - Systems which deal with a lot of data have a lot of work like uploading data and computing on them rather than retrieving it.