images, and videos shared by users worldwide. This volume is continuously growing, requiring scalable storage and processing solutions.

2. Velocity:

- **Definition:** Refers to the speed at which data is generated, processed, and analyzed.
- **In Social Media:** The data flow on social media is extremely rapid. Posts, likes, comments, shares, and reactions happen in real time. Social media platforms must handle these high-velocity interactions to ensure a smooth user experience, deliver real-time recommendations, and monitor trends or viral content.

3. Variety:

- **Definition:** Refers to the different types of data formats being generated.
- **In Social Media:** Social media platforms deal with diverse data types, including structured data (user profiles), unstructured data (images, videos, text posts), and semi-structured data (hashtags, emojis, and metadata). Handling and analyzing these various formats is a challenge, but it provides rich insights into user behavior and trends.

4. Veracity:

- **Definition:** Refers to the quality, accuracy, and trustworthiness of the data.
- **In Social Media:** Social media data often comes with challenges of trust and accuracy. User-generated content may include spam, fake accounts, or misinformation. Platforms must constantly verify the authenticity of data, ensure content is reliable, and filter out noise, like bots or irrelevant posts, to extract meaningful insights.

These characteristics illustrate the complexity of managing big data on social media platforms and emphasize the importance of effective data processing and analytics strategies.

Q3. Explain how node failure is handle in Hadoop?
→
Hadoop handles node failures through the following mechanisms:

1. **DataNode Failure:**
    - Data is replicated across multiple DataNodes (default 3 replicas). If a DataNode fails, the NameNode detects the failure and re-replicates the data to maintain the required number of copies.
    - Clients can read data from other replicas on different nodes.
2. **NameNode Failure:**
    - In **High Availability (HA)** setups, two NameNodes exist: Active and Standby. If the Active NameNode fails, the Standby automatically takes over, ensuring uninterrupted operation.
    - Metadata is backed up using **FsImage** and **EditLogs**.
3. **Task Failure:**
    - Failed tasks are rescheduled on different nodes by the JobTracker (or ResourceManager in YARN) to ensure job completion without disruption.