

Data Lifecycle Scenario: Social Media Analytics

You are given a short scenario to analyze.

Step 1: Individual work Draw the lifecycle diagram, indicating:

- the sources of data,
- potential issues with quality, trustworthiness, or usability,
- the expected type(s) of data (structured / semi-structured / unstructured)

Step 2: Group Discussion: Compare your individual diagrams and notes with group members. Reflect on each stage of the lifecycle with your team and answer the guiding questions provided, completing the **lifecycle-step boxes** together. Prepare a brief summary (keywords or 1–2 sentences) to present.

Time guideline: 5 minutes individually, 10 minutes in groups.

Scenario Description: Social Media Analytics

A social media platform collects posts, likes, shares, comments, and event RSVPs, generating large volumes of unstructured data. The platform uses this data to monitor trends, detect spam or bot activity, and improve content recommendations. Data is collected continuously and includes text, images, videos, and metadata. Analysts must handle high-velocity data streams, ensure user privacy, and detect patterns or anomalies in real time. Social and ethical considerations must also be accounted for, such as moderation policies and bias in algorithmic recommendations.

Scenario Analysis (individual work): Draw the lifecycle diagram, indicating:

- Source(s) of the data (human/machine, internal/external)
- Data type (structured / semi-structured / unstructured)
- Any obvious quality, reliability, or trustworthiness issues

Lifecycle diagram

Data Lifecycle Reflection (group work): For each stage, reflect on what it could involve in this scenario. Consider:

- How to filter out spam, bots, or irrelevant content?
- How to handle very large volumes of unstructured data?
- How can insights be communicated to marketing or product teams?
- What ethical considerations arise when analyzing user content?

Collect

Clean / Preprocess

Store

Analyze

Communicate / Report