Top-10 most viewed video

1. What are the names and NetIDs of all your team members? Who is the captain? The captain will have more administrative duties than team members.

   Captain: You Zhou - youzhou3
   Team members: Tianyi Zhang - tianyi18
   Tianyu Zhou - aliang7
   Shuxian Xu - sx30

2. What is your free topic? Please give a detailed description. What is the task? Why is it important or interesting? What is your planned approach? What tools, systems or datasets are involved? What is the expected outcome? How are you going to evaluate your work?

   Topic: Top 10 most viewed videos
   Description:
   The primary goal of our project is to identify and analyze the top 10 most viewed videos across a simulating platform.
   Task:
   - Data feed
   - Data parsing
   - Platform construction
   - Map-reduce
   - Result rendering

   Tools:
   - Python libraries pandas, numpy
   - Docker, Nginx, Gunicorn for network orchestration
   - Caching

   Why:
   With the explosion of digital content, platforms are inundated with large volumes of data. Recognizing what captivates an audience in this sea of content can provide an edge. For our simulating platform, insights into top-performing videos can steer content creators, enhance platform features, and optimize content recommendations.
   Outcome:
   A robust, scalable system that, when provided with raw video data, can autonomously process, analyze, and produce a list of the top 10 most viewed videos. This system will serve as a valuable tool for content strategists, platform operators, and advertisers.
   Evaluation:
   Efficiency: The system's capability to process data and generate results swiftly.

Accuracy: Ensuring that the derived top 10 videos genuinely reflect the highest view counts.
Scalability: The system's ability to handle larger datasets as the platform grows without significant modifications.

3. Which programming language do you plan to use?

   Python

4. Please justify that the workload of your topic is at least 20*N hours, N being the total number of students in your team. You may list the main tasks to be completed, and the estimated time cost for each task.
   a. Data feed - 10hr
   b. Data parsing - 5hr
   c. Platform construction - 20hr
   d. Map-reduce - 35hr
   e. Result rendering -10hr