**1. Create and submit a problem statement from your proposal.**

On Social Media Platform users are uploading number of images, there is high possibility of duplicate images can be uploaded by different users. It’s doesn’t make any sense to store same images in to storage which is going to reserved unnecessary storage. We should some short of logic to that can identify the duplicate images and review from storage.

**2. Felicitate functional and non-functional requirements for your project.**

**Non-Functional :**

1) collection of Images, some duplicate images

2) Installation of Linux Virtual machine

3) Installed the Hadoop Environment, create 1 Master and 2 Slave Node

4) Configuration of Hadoop Environment, JAVA and eco system

**Functional :**

1) Create a map Reduce Program using algorithm to read the files from HDFS. Convert the files into Sequence file format. Compare the files to identify the duplicate images and remove from HDFS.

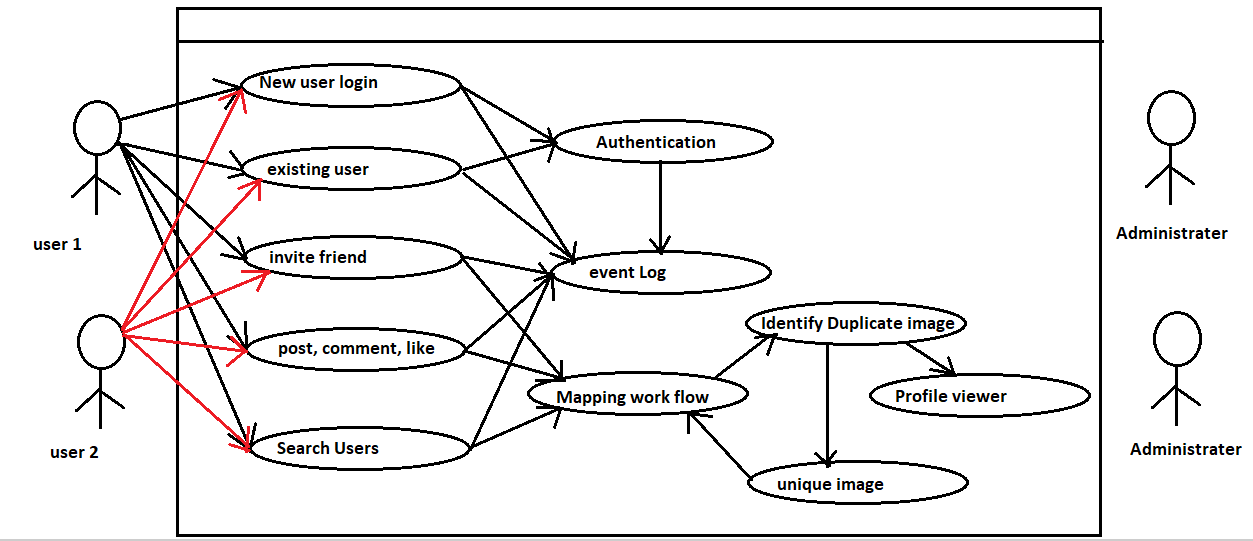
2) Set up HDFS directory

3) Execute the Map Reduce code on Hadoop Environment

4) Schedule of Job, set up monitor and error handling Process

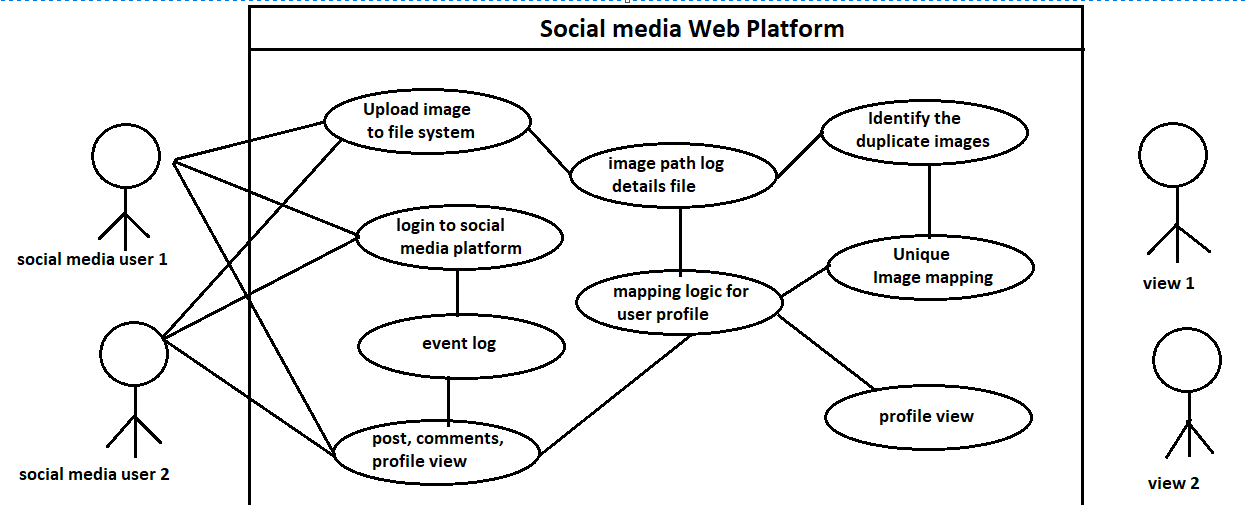
**3. Develop a Software Requirements Specification (SRS) and use cases (user stories - define the main flow, alternate flow and exception flow for use cases/scenarios)**

**Facebook social web platform:**

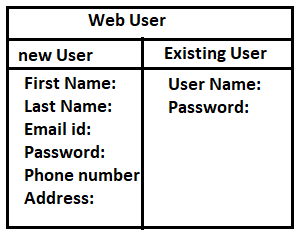


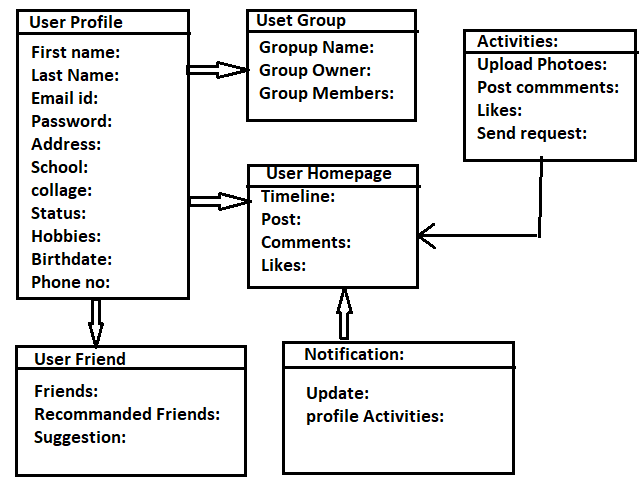
**4. Submit use cases, class and sequence diagrams**

**1. use case Diagram:**



**2. class Diagram:**





**3. Sequence Diagram:**

