

SPA Assignment – July-2021

A startup company wants to build a private community app which will allow People to create a specific community and invite other users to join that community. Each community in itself is a social network like Facebook specifically for the people in that community.

Users can mainly have two workflows where, A community user can

1. do various types of posts (including text, images, videos) in community that will show up in feed of other users, interact with the posts of other users (like, comment, share etc.).
2. User can do conversation with other users via Direct messaging or there can be various user discussion groups within that community where multiple users can do conversations.

In an active community of million users the load on systems will increase a lot and traditional RDBMS techniques will not be able to handle the loads. You have been appointed to design a scalable solution for the same using Big Data Technologies with the help of the system can efficiently scale to millions/billions of users and fulfill their requests in an efficient way. Additionally, getting insights/analytics from such an app is also another use case for the organization.

You can make the appropriate assumptions for the events data, the analysis that needs to be done on it. Make sure all those details are also made available while submitting the data solution.

Submission Requirements

Part 1

Note: Make necessary assumptions regarding the users creating different data points like posts, messages, likes, comments, shares etc.

1. Propose a suitable architecture of the system with high level details of the components involved in building the complete system. [4 Marks]
2. Detailed document on the structure of data you would store, from data flow tier to master storage to delivery tier and how the interaction between different subcomponents of the system work. [4 Marks]
3. Link of a short 5–10-minute video explaining the above submissions.
Note. You may host the video on a google drive/youtube/anywhere and share the publicly accessible link of the video in a separate Doc/PDF file. [2 Marks]

Part 2

Note: Make necessary assumptions/generate fake data, regarding the users creating different data points like posts, messages, likes, comments, shares etc.

1. Out of the two main workflows defined in top

- a. Social Media Posts complete workflow
- b. Messenger/Conversations between users/user groups

Choose any one workflow and implement it using open source Technologies such as Kafka, Spark, Flink etc. and in one programming language Python/Java. [6 Marks]

2. Create a streaming analytics pipeline and a dashboard that shows Realtime insights of the application.

Note: Based on your workflow decide what could be valuable data points you can gather and generate insights out of it. [6 Marks]

3. Submit both the codes

- a. Of the working project
- b. Of the analytics pipeline

Separately and Link of a short 5–10-minute video helping to understand how the integration between the different system subcomponent works. Proper flow needs to be shown between the different classes defined for the workflow and data pipeline. [3 Marks for the Video]