(MS Deep Singh)

Video 36. StackOverflow / Quora System Design.

↳ {LeetCode can also be designed like this}

Functional Requirements

1. Post Questions
   - (a) User should be logged in to post a question.
   - (b) associate question ~~must have~~ can add tag(s).
   - (c) upvote / downvote can be done for question.
   - (d) add comments to questions → If some comments is posted, we cannot add reply to that comment.
   - (e) Question can contain text/images.
2. Post Answer.
   - (a) User should be logged In
   - (b) Multiple answers can be there for a question
   - (c) upvote / downvote answers [Same as post questions]
   - (d) add comments to answer.
   - (e) Single user can write single answer [If required he will edit the same answer.]

③ questions Newsfeed → If gone to home page of stackoverflow we can see newsfeed of questions

④ User should be notified for questions they have subscribed to.

⑤ Search for Questions → { Using keywords }

⑥ User time line → Questions / answers / actions → On Profile Shewing all activities

PI requirements → { We will not be discussing }.
But our solution will be able to extend it

① Rank answers for a question.

② views for a question / answer.

[Non-Functional Requirements]

① Availability

② Latency should be low

③ Eventual Consistency is fine

④ Questions and answers written should be persisted, unless deleted.
↳ Durable

(Capacity Estimations)

Traffic: 270 Million visits/month.

$$TPS = \frac{270,000,000}{30 \times 24 \times 60 \times 60} = 104 \quad (\text{Read TPS} = 110/s)$$

Assumptions: Lets say out of all users.

20% of users ask questions → 20 ques/sec

20% of users answer the questions → 20 answers/sec

Approx Size of 1 ques/answer = 100KB. = 40 × 100 = 4000KB/sec

= 4MB/second.

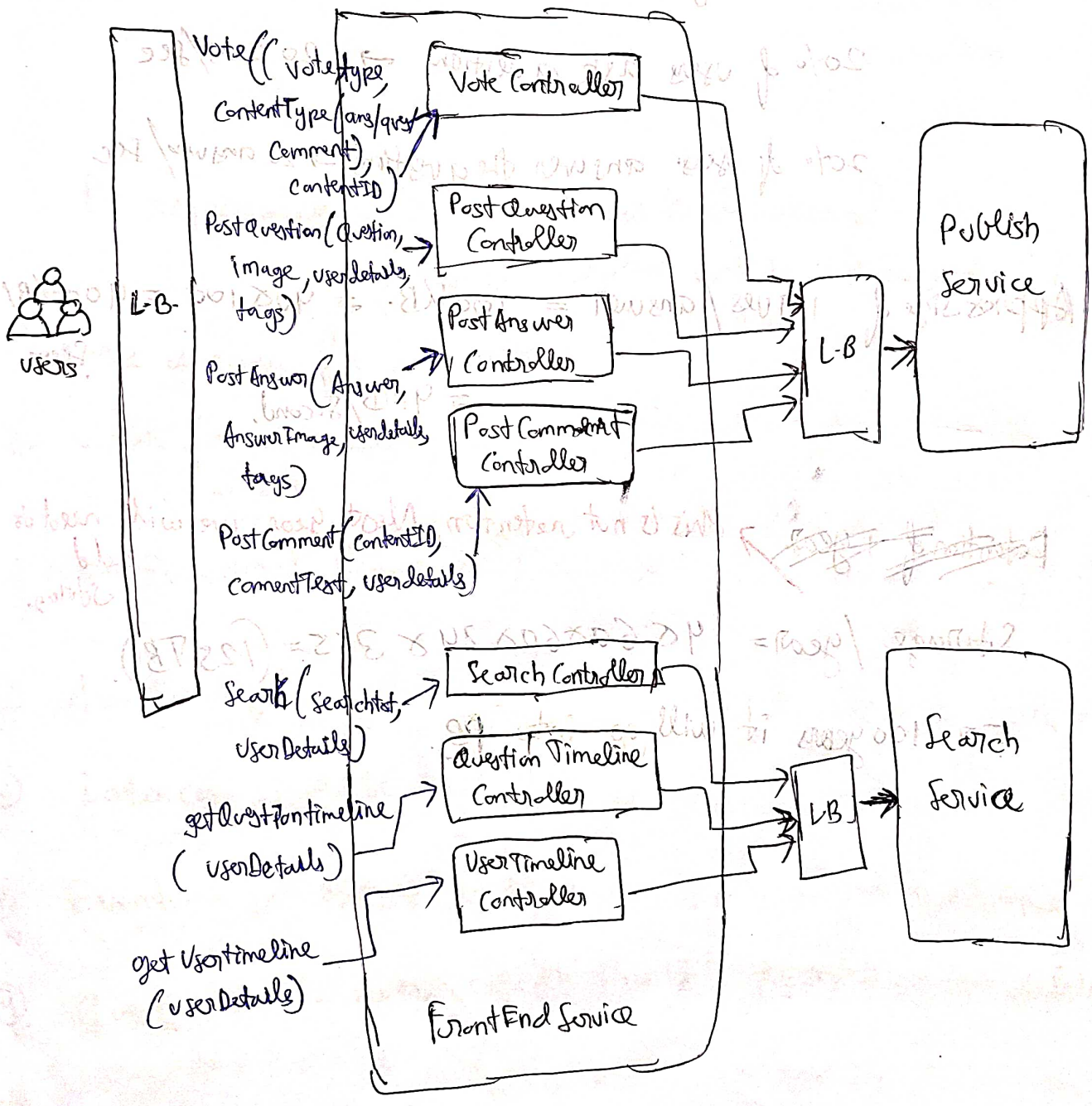
~~Retention of 1 year~~ → This is not retention, Next year we will need to add storage.

Storage/year = 4 × 60 × 60 × 24 × 365 = (125TB).

For 100 years it will go into PB.

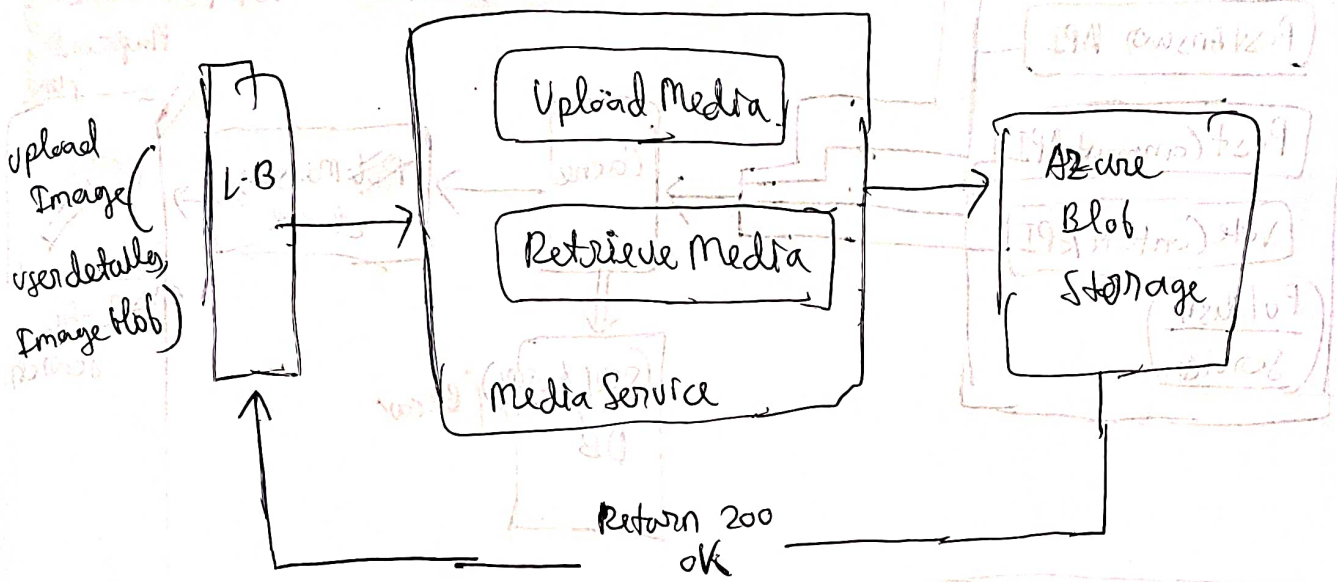
High level Architecture [If we explicitly talk about Stack Overflow, It is built on microsoft stack, Like Azure blob Storage and all].

- Question is parent entity
- multiple answers can be associated to a question.
- multiple comments can be associated to a question/answer.

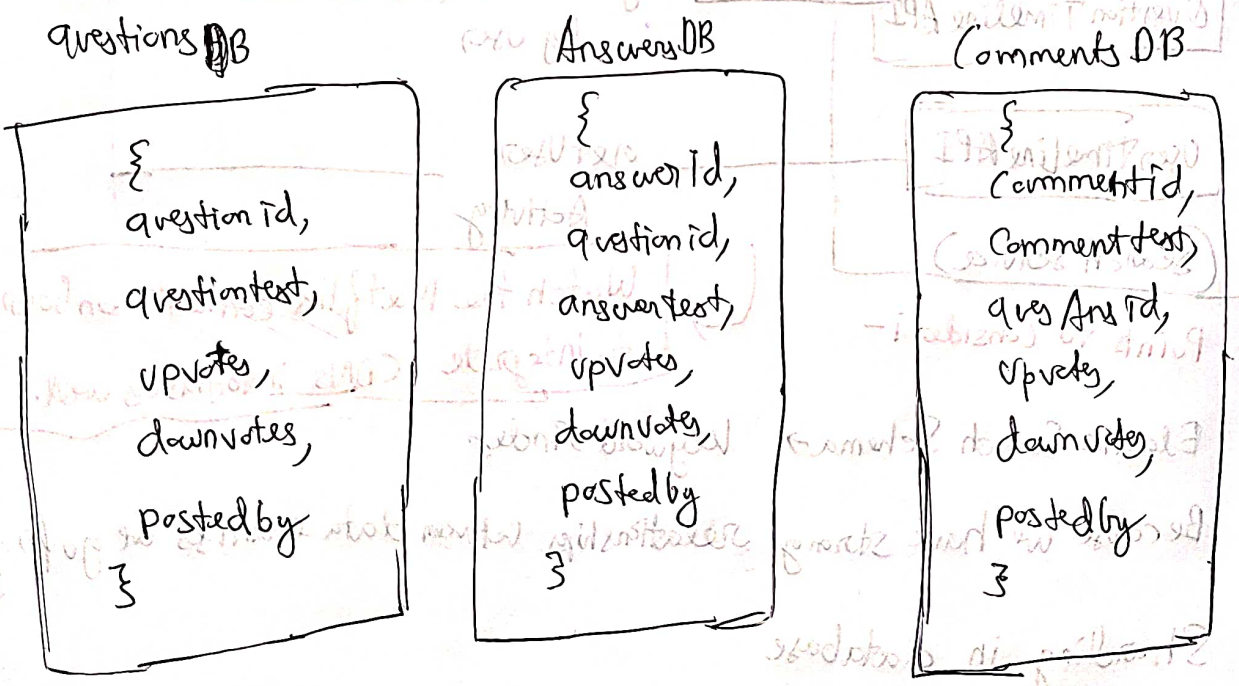


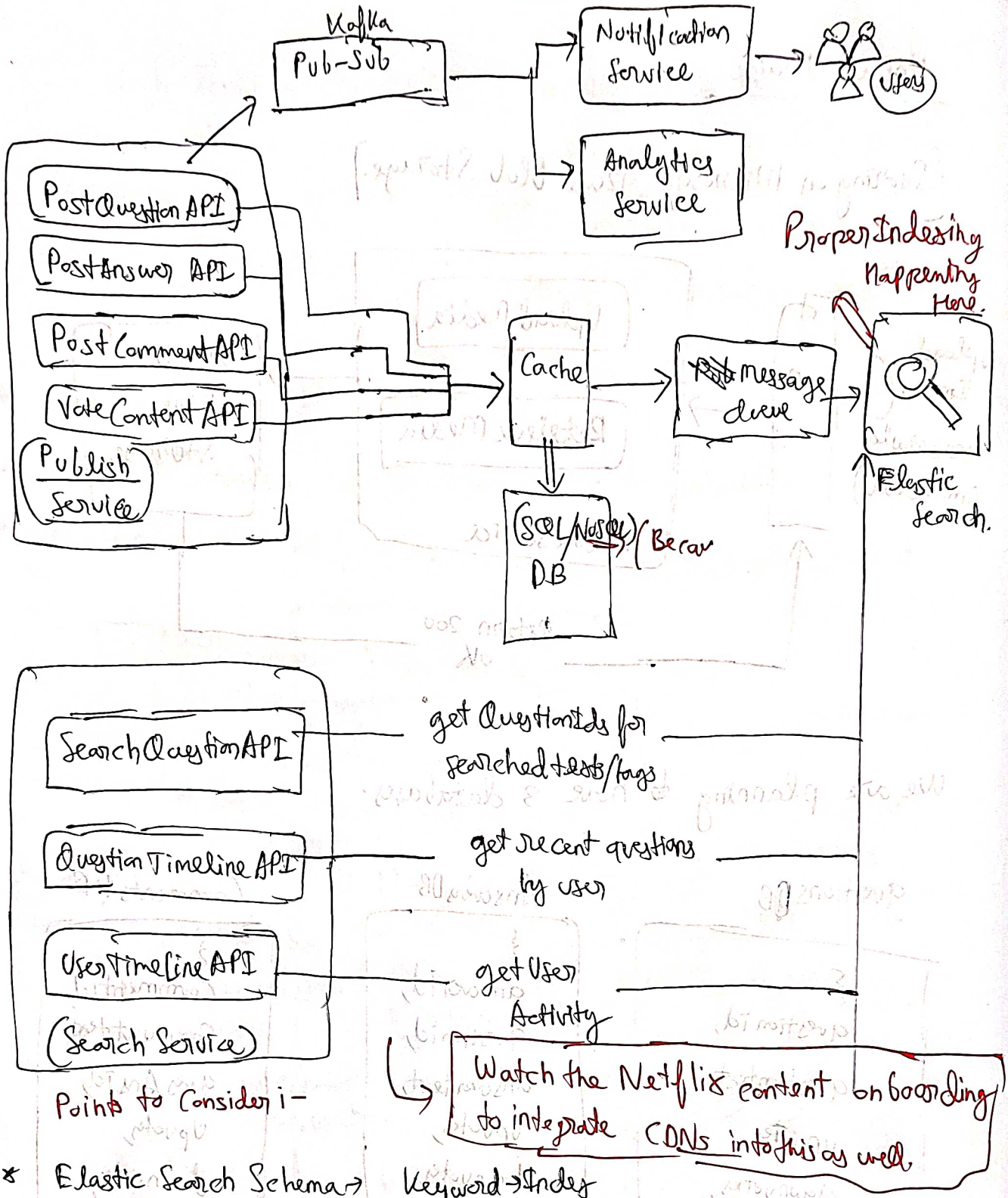
Upload Image

[Storing on Microsoft Azure Blob Storage.]



We are planning to have 3 databases.





Points to Consider i-

- Elastic Search Schema → Keyword → Index
- Because we have strong relationships between data entities we go for RDS/SQL.
- Sharding in database
- we can make PostAnswer, PostQuestion, UpVotes, Downvotes as async
- Background Scheduler job to update user badges.
- Caching requirement and Strategies