

we want to design stackoverflow

scope and functional requirements

post_question(string)

post_answers(question_id, answer)

post_comment(answer_id, comment)

like_comment(comment_id)

view_question → List[answers], for each answer comments, for each comment, likes

search_question(text)

upvote(answer_id)

performance requirements

3M DAU

load

questions posted per day → 1

questions viewed per day → 5 ⇒ read heavy system

answers posted per day → 1

1 → 1MB

cap plan: $0.1 * 3M * 1MB * 400 \text{ days} * 10\text{years} = 4e3 * .1 * 3M * 1MB = 1 * 10^9 \text{ MB} = 1\text{PB} \leftarrow \text{storage}$

data model

question_id, question_str, media_id (cloudfront + S3),

answer

answer_id, question_id, answer_votes, media_id

comment

id, answer_id, votes

high level design

relational database → synchronous replication

pagination → pre-cache cache next page