HLD Quora

1) Define MVP → Minimum Viable Product
2) Estimation of Scale → Storage, Read/Write, QPS (Queries per sec)
3) Design Goals → latency, Availability/consistency
4) API & Design Challenges

---

1) Define MVP

a) Post Questions
b) Answer Questions
c) Upvote/Downvote
d) User Profile → login/logout
e) News feed ✓
f) Searching ← (not in MVP)
g) Topics for a question (user can follow topics)
h) Ranking Answers.

---

2) Estimation of Scale

≈ 300 M MAU (Monthly Active Users)
≈ 50 M DAU (Daily Active Users)
≈ 10K questions per day
≈ 5 Answers per question

a) Data to Store →

Questions
Answers
> heavy storage

Tags/Topics
User Data
> less data

5years

Questions → 10K que/day

0.5KB

$10^4 * 365 * 5 \longrightarrow$ 18.25M   (1M → 10^6)

Answers → 5 per question

2KB

$18.25 * 5 \longrightarrow$ 91.25M

Q/A → text → string → set of characters → 1B

Question → 0.5KB   (500 characters)

Answer → 2KB   (2000 characters)

Data to store → $(18.25 * 0.5 \; B + 91.25 * 2 \; B) * \overset{M}{10^6} * \overset{K}{10^3}$

$\approx 190 * 10^9 \longrightarrow$ 190 GB

200GB → store in one machine

Replication → store same data in multiple machines.

Read / Write Heavy

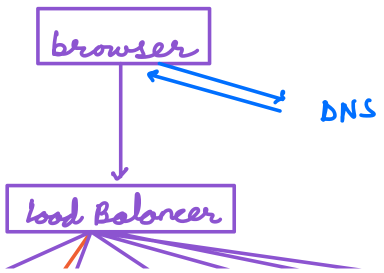
50 M   DAU

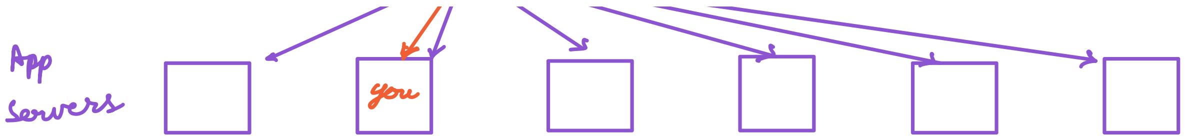
10K questions per day

Read Heavy System

5 Answers per question

QPS → Queries Per Second





Design Goals 1) latency → important but not critical.

2) Consistency vs Availability (CAP)

X ✓

Same data at all times.

Storage → 200GB (1 Machine)

MySQL

Tables → Questions, Answers, Topics, User Data

(time when que was added)

Questions → id, userId, text, t.s, upvotes, downvotes)

Answers → id, userId, text, t.s, qId, upvotes, downvotes)

Topics → id, text, t.s

User Data → id, name, email, ...

Question_Topic → qId, tId

User Topic → uId, tId

API → 1) add Questions (user_Id, text, topics, queId, ts)

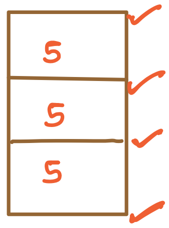
just one entry in question table.

2) add Answer (user_Id, text, queId, ansId, ts)

just one entry in answer table.

3) upvoteQuestion (userId , questionId) }
4) downvoteQuestion (userId , questionId) } same for answer

5) getNewsFeed (userId , numberOfResults , offset)
(homepage) 5 / 10 / 20 ✓ 1/6/11



a) User without log-in

✓ popular questions → asked recently
most upvote questions ✓

Question → 100 votes ✓ (today)
Question → 1000 votes ✓ (last week)

Time Decay → 1.1 per day (subjective)

110 → $110/1.1 = \boxed{100}$ → $100/1.1 = 91$ . . .
today day 2 day 3

Question1 → 60 upvotes + 50 upvotes + 100 upvotes
today yesterday day before yesterday

$\frac{50}{1.1} = 45$ $\frac{100}{(1.1)^2} = 82$

$60+45+82 = 187$ ✓

Question 2 → 100 upvotes today ✓

Decay
1.2 per day 100 today 120 yesterday 150 day before
100 $120/1.2 = 100$ $150/(1.2)^2 = 104$ ✓

log-in user

user → topics → question wrt topics

limited

cache (store) ✓

topic → top 20 questions (recency + upvotes)

→ update after a certain time (after every 4 hours ← eg) ✓

→ when for a question the count of upvotes reach multiple of 50 or 100. ✓

H.W → Rank answers for a question. ✓