Design Facebook Newsfeed

Wednesday, March 27, 2024   8:50 PM

similar ones - Twitter, LinkedIn, Instagram, etc

FUNCTIONAL REQUIREMENTS

1. create all factor/posts made by people, groups, pages that a user follow
2. Favorite or preferred posts
3. Posts may include text, photos and videos
4. Add new posts to the news feed.

NON FUNCTIONAL REQUIREMENTS

1. Availability over consistency
2. Low latency

CAPACITY ESTIMATION

Total users 1.5 billion.
  300 M   active users (daily)
User is going to fetch timeline 5 times a day.

  300 friends , 200 pages

Traffic: 300 M × 5 ≈ 1500 × 10⁶ ≈ 1.5 Billion newsfeed req/day

$$300 M \times 5 \approx 1500 \times 10^6 \approx 1.5 \text{ Billion newsfeed req/day}$$

Storage:
  Newsfeed has 500 posts
  Each post → 1kb
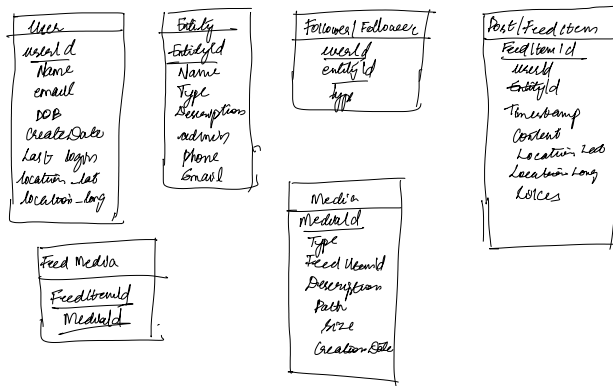 Lets say 500 kb data / user

$$1.5 \times 10^9 \times 500 \times 1kb = 750 \times 10^9 \times 10^3 = 750 \text{ TB/ day}$$

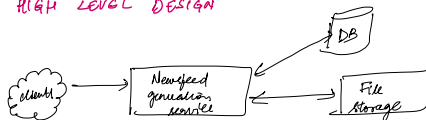$$1.5 \times 10^9 \times 100 \times 1Kb = 150 \text{ PB/day.}$$

APIs

1. List of   ← getNewsfeed (api-dev-key, user_id, preference-list, feed items   post count)

DB

User
userId
Name
email
DOB
CreateDate
Last login
location_lat
location_long

Entity
EntityId
Name
Type
Description
admin?
Phone
Email

Follower/Followee
userId
entityId
Type

Post/FeedItem
FeedItemId
userId
EntityId
Timestamp
Content
Location Lat
Location Long
Likes

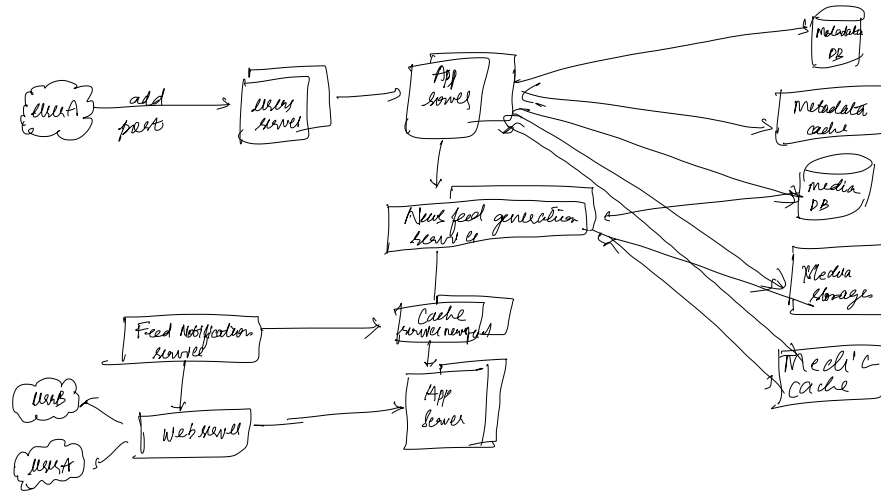| Media |
| --- |
| MediaId |
| Type |
| FeedItemId |
| Description |
| Path |
| Size |
| Creation Date |

Feed Media
FeedItemId
MediaId

HIGH LEVEL DESIGN

LOW LEVEL DESIGN

HashMap
Key : Value
userId : List of posts

Linked HashMap
FeedItemId : FeedItem, DateTime, Last generated.

DATA PARTITIONING
Posts and metadata:

userId: Hot user problem
Post Id: Hot post problem
Post creation time : Load on server
Post Id + creation time
(epoch + seqnumber)   best choice
                         ↓
           Consistent hashing

SHARDING OF NEWSFEED
memory
userId? ✓

RANKING
Time of update
Likes
social graph
depends on content favorites
video

HOW THE NEWS FEED WORKS

1. Inventory   2. Signals
                       Likes, Comments

3. Prediction   – which of posts has videos or photos
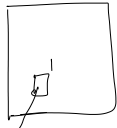4. Scoring   – cumulative score for each feedId

Storing and collecting all information

Push ⟶ more load on server.
Pull ⟶ more reqs

✓ Hybrid ⟶ Push + Pull
            (others)   (Celebrity)

scroll to top
for new newsfeed