System Design Parameters

What would be a good system design (high level) to handle incoming messages?

- The system would need to handle periods with high bursts of messages (10k m/s) and periods with no messages at all (during the night for instance).
- Messages are collected from multiple social media sources at different (random) intervals and are matched to cases.
- We discard the messages that don't match to any case.
- Different users might see the same message if they have similar queries.
- Messages are delivered in realtime to the frontend.
- Also important to mention that we have multiple data enrichment pipelines like **language recognition**, **text summarization**, **sentiment analysis,** etc. We expect to see a high level diagram explaining the flow of data.
- No need to expand too much on the data structures.
- Feel free to make your ssumptions, there's no single right answer. We will value the thought process more than the end result.

Assumptions :

- DAU : 1 million
- Request Size : 0.3 MB
- Average Request Size : 100 Bytes
- Time To Store : 10 years
- Requests Per Day : 200 Million

Traffic Estimation

$60\ seconds\ x\ 60\ minutes\ =\ 3600\ secs\ per\ hour$
$3600\ x\ 24\ hours\ =\ 86400\ secs\ per\ day$
$1\ Million\ DAU\ x\ 200\ queries\ =\ 200\ Million\ Query\ Requests$
$200\ Million\ Requests\ //\ 86400\ =\ 2315\ Requests\ per\ second$

Memory Estimation

$200\ Million\ Requests\ x\ 100\ Bytes\ =\ 20GB$
$20\ GB\ x\ 0.2\ =\ 4GB$
$4GB\ x\ 3\ =\ 12\ GB$

Bandwidth Estimation

$200\ Million\ Requests\ x\ 0.3\ MB\ =\ 60000GB$
$60000\ //\ 86400\ =\ 0.69\ GB\ per\ second$

Storage Estimation

$1\ Million\ Requests\ x\ 0.3\ MB\ =\ 0.3\ TB\ per\ day$
$0.3\ TB\ x\ 365\ days\ x\ 10\ years\ =\ 1.1\ Petabytes$