Requirements of Quora's Design

Learn about the requirements for designing Quora.

We'll cover the following

- Requirements
  - Functional requirements
  - Non-functional requirements
- Resource estimation
  - Number of servers estimation
  - Storage estimation
  - Bandwidth estimation
- Building blocks we will use

Requirements

Let’s understand the functional and non-functional requirements below:

Functional requirements

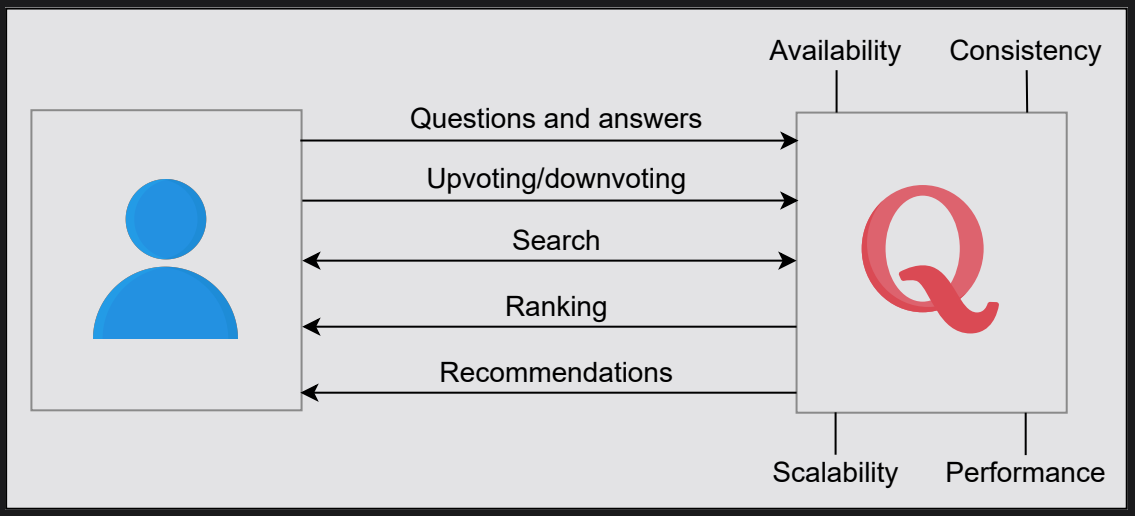
A user should be able to perform the following functionalities:

- **Questions and answers**: Users can ask questions and give answers. Questions and answers can include images and videos.
- **Upvote/downvote and comment**: It is possible for users to upvote, downvote, and comment on answers.
- **Search**: Users should have a search feature to find questions already asked on the platform by other users.

- **Recommendation system**: A user can view their feed, which includes topics they're interested in. The feed can also include questions that need answers or answers that interest the reader. The system should facilitate user discovery with a recommender system.
- **Ranking answers**: We enhance user experience by ranking answers according to their usefulness. The most helpful answer will be ranked highest and listed at the top.

Non-functional requirements

- **Scalability**: The system should scale well as the number of features and users grow with time. It means that the performance and usability should not be impacted by an increasing number of users.
- **Consistency**: The design should ensure that different users' views of the same content should be consistent. In particular, critical content like questions and answers should be the same for any collection of viewers. However, it is not necessary that all users of Quora see a newly posted question, answer, or comment right away.
- **Availability**: The system should have high availability. This applies to cases where servers receive a large number of concurrent requests.
- **Performance**: The system should provide a smooth experience to the user without a noticeable delay.



Functional and non-functional requirements of Quora

Resource estimation

In this section, we'll make an estimate about the resource requirements for Quora service. We'll make assumptions to get a practical and tractable estimate. We'll estimate the number of servers, the storage, and the bandwidth required to facilitate a large number of users.

Assumptions: It is important to base our estimation on some underlying assumptions. We, therefore, assume the following:

- There are a total of 1 billion users, out of which 300 million are daily active users.
- Assume 15% of questions have an image, and 5% of questions have a video embedded in them. A question cannot have both at the same time.
- We'll assume an image is estimated to be 250 KBs, and a video is considered 5 MBs.

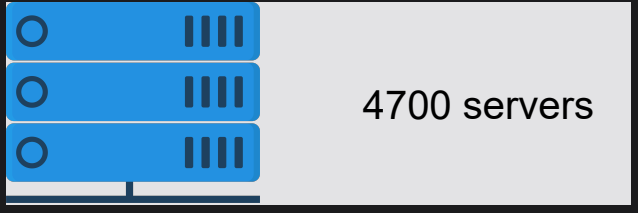
Number of servers estimation

Since we have 300 million daily active users for Quora. Considering our assumption of using daily active users as a proxy for the number of requests per second to find the number of servers for peak load times, we get 300 million requests per second. Then, we use the following formula to calculate the number of servers:

$$Servers\ needed\ at\ peak\ load = \frac{Number\ of\ requests/second}{RPS\ of\ server}$$

Using 64,000 as an estimated RPS a server can handle, the required servers are estimated as follows:

$$Servers\ needed\ at\ peak\ load = \frac{300\ million}{64,000} = 4687.5 \approx 4.7K\ servers$$



The estimated number of servers required for Quora

Storage estimation

Let’s keep in mind our assumption that 15% of questions have images and 5% have videos. So, we’ll make the following assumptions to estimate the storage requirements for our design:

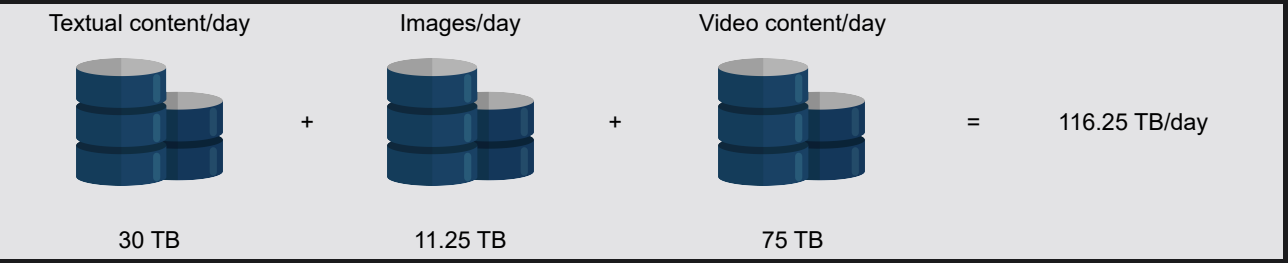
- Each of the 300 million active users posts 1 question in a day, and each question has 2 responses on average, 10 upvotes, and 5 comments in total.
- The collective storage required for the textual content (including the question, answer(s), and comment(s) text) of one question equals $100\ KB$.

Storage Requirements Estimation Calculator

Questions per user	1	per day
Total questions per day	300	millions
Size of textual content per question	100	KB
Image size	250	KB
Video size	5	MB
Questions containing images	15	percent
Questions containing videos	5	percent

Storage for textual content	f 30	TB
Storage for image content	f 11.25	TB
Storage for video content	f 75	TB

See Detailed Calculations



Summarizing storage requirements of Quora

> Total storage required for one day =
> $30\ TB + 11.25\ TB + 75\ TB\ = 116.25\ TB$ per day

The daily storage requirements of Quora seem very high. But for service with 300 million DAU, a yearly requirement of $116.25\ TB \times 365 = 42.43\ PB$ is feasible. The practical requirement will be much higher because we have disregarded the storage required for a number of things. For example, non-active (out of $1\ B$) users' data will require storage.

Bandwidth estimation

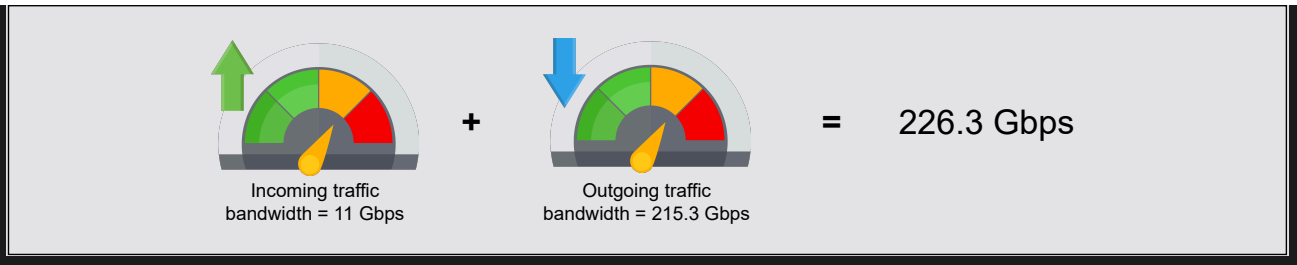
The bandwidth estimate requires the calculation of incoming and outgoing data through the network.

- **Incoming traffic**: The incoming traffic bandwidth required per day will be equal to $\frac{116.25\ TB}{86400} \times 8 = 10.9\ Gbps \approx 11\ Gbps$
- **Outgoing traffic**: We have assumed that 300 million active users views 20 questions per day, so the total bandwidth requirements can be found in the below calculator:

Bandwidth Requirements Estimation Calculator

Total storage required per day	116.25	TB
Incoming traffic bandwidth	*f* 11	Gbps
Questions viewed per user	20	per day
Total questions viewed	*f* 69444	per second
Bandwidth for text of all questions	*f* 55.56	Gbps
Bandwidth for 15% of image content	*f* 20.83	Gbps
Bandwidth for 5% of video content	*f* 138.89	Gbps
Outgoing traffic bandwidth	*f* 215.3	Gbps

Detailed Calculations



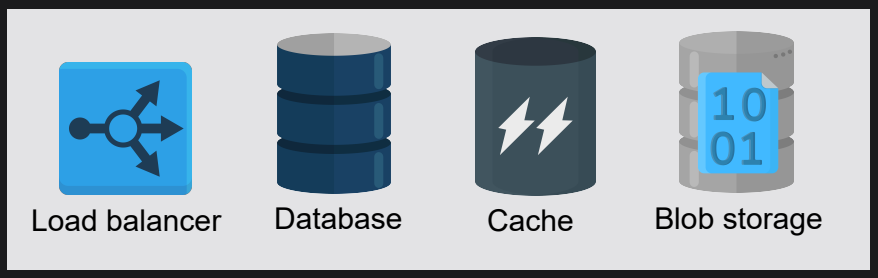
Summarizing the bandwidth requirements of Quora

Total bandwidth requirement of Quora is equal to:

> Incoming + outgoing traffic bandwidth
> $= 11\ Gbps + 215.3\ Gbps = 226.3\ Gbps$

Building blocks we will use

We'll use the following building blocks for the initial design of Quora:



Building blocks required for our design

- **Load balancers** will be used to divide the traffic load among the service hosts.
- **Databases** are essential for storing all sorts of data, such as user questions and answers, comments, and likes and dislikes. Also, user data will be stored in the databases. We may use different types of databases to store different data.
- **A distributed caching system** will be used to store frequently accessed data. We can also use caching to store our view counters for different questions.
- **The blob store** will keep images and video files.