



Requirements of Google Docs' Design

Learn about the requirements for designing a collaborative editing service.

We'll cover the following



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

Requirements

Let's look at the functional and non-functional requirements for designing a collaborative editing service.

Functional requirements

The activities a user will be able to perform using our collaborative document editing service are listed below:

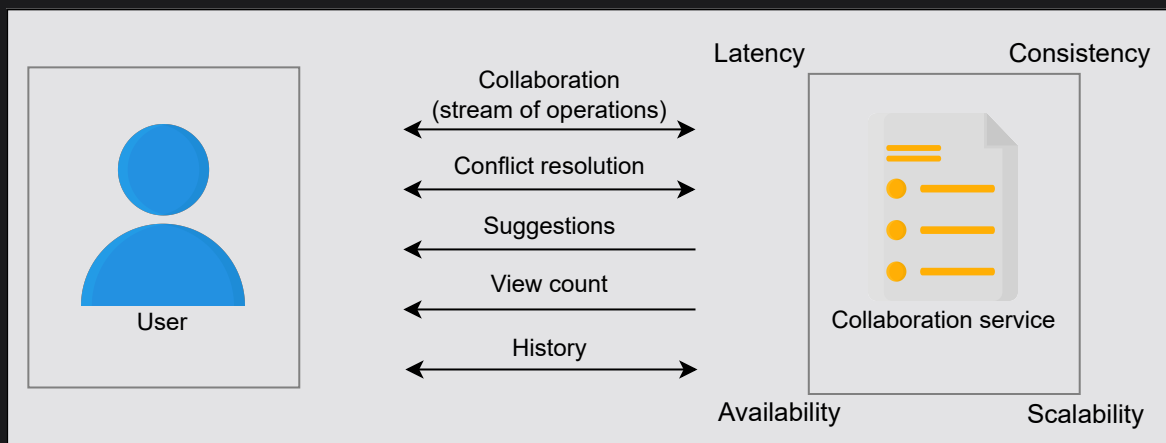
- **Document collaboration:** Multiple users should be able to edit a document simultaneously. Also, a large number of users should be able to view a document.
- **Conflict resolution:** The system should push the edits done by one user all the other collaborators. The system should also resolve conflicts



between users if they're editing the same portion of the document.

- **Suggestions:** The user should get suggestions about completing frequently used words, phrases, and keywords in a document, as well as suggestions about fixing grammatical mistakes.
- **View count:** Editors of the document should be able to see the view count of the document.
- **History:** The user should be able to see the history of collaboration on the document.

A real-world document editor also has to have functions like document creation, deletion, and managing user access. We focus on the core functionalities listed above, but we also discuss the possibility of other functionalities in the lessons ahead.



Functional and non-functional requirements of collaborative editing service

Non-functional Requirements

- **Latency:** Different users can be connected to collaborate on the same document. Maintaining low latency is challenging for users connected from different regions.
- **Consistency:** The system should be able to resolve conflicts between users editing the document concurrently, thereby enabling a consistent view of the document. At the same time, users in different regions should see the updated state of the document. Maintaining consistency is important for users connected to both the same and different zones.
- **Availability:** The service should be available at all times and show robustness against failures.



- **Scalability:** A large number of users should be able to use the service at the same time. They can either view the same document or create new documents.

>

Resource estimation

Let's make some resource estimations based on the following assumptions:

- We assume that there are 80 million daily active users (DAU).
- The maximum number of users able to edit a document concurrently is 20.
- The size of a textual document is 100 KB.
- 30% of all the documents contain images, whereas only 2% of documents contain videos.
- The collective storage required by images in a document is 800 KB, whereas each video is 3 MB.
- A user creates one document in one day.




Based on these assumptions, we'll make the following estimations.

Storage estimation

Considering that each user is able to create one document a day, there are a total of 80 million documents created each day. Below, we estimate the storage required for one day:

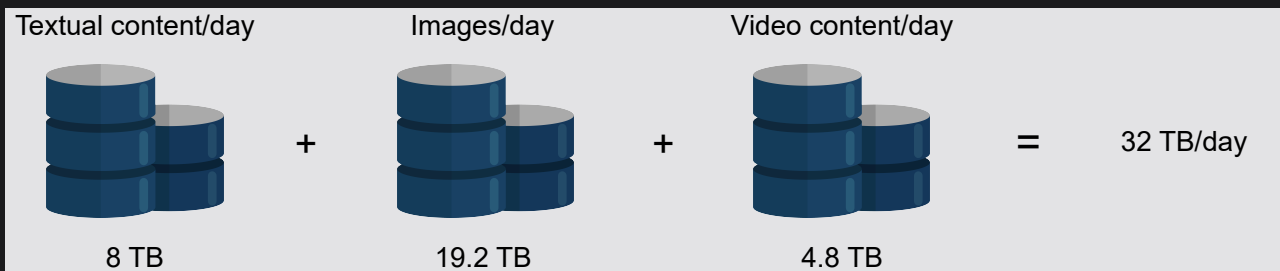
Note: We can adjust the values in the table below to see how the storage requirement estimations change.

Estimation for Storage Requirements

Number of documents created by each user	1	per day	 
Number of active user	80	Million	

s		
Number of documents in a day	^f 80	Million
Storage required for textual content per day	^f 8	TB
Storage required for images per day	^f 19.2	TB
Storage required for video content per day	^f 4.8	TB
Total storage required per day	^f 32	TB

 [See Detailed Calculations](#)



Storage required by online collaborative document editing service per day

Total storage required for one day is as follows: $8 + 19.2 + 4.8 = 32 \text{ TBs}$ per day

Note: Although our functional requirements state that we should keep a history of documents, we didn't include storage requirements for historical data for the sake of brevity.



Bandwidth estimation

Incoming traffic: Assuming that 32 TB of data are uploaded per day to the network of a collaborative editing service, the network requirement for incoming traffic will be the following:

$$\frac{32\text{ TB}}{86400} \times 8 = 3\text{Gbps approximately}$$

Outgoing traffic: To estimate outgoing traffic bandwidth, we'll assume the number of documents viewed by a user each day. Let's consider that a typical user views five documents per day. Then, the following calculations apply:

Note: We can adjust the values in the table below to see how the calculations change.

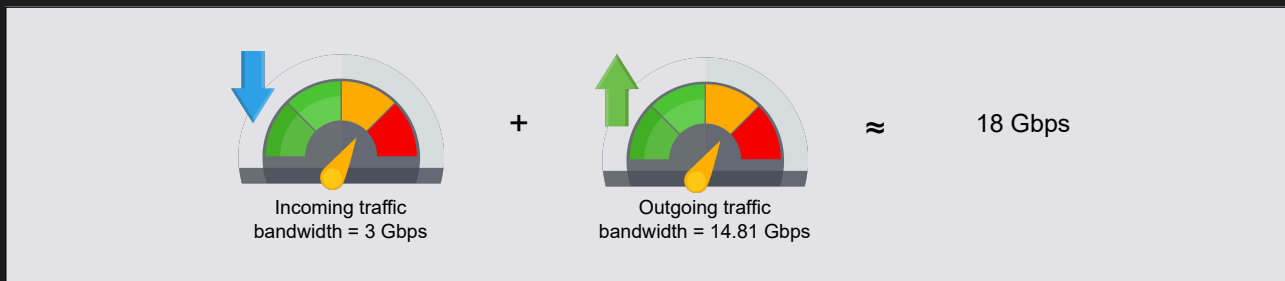
Number of documents viewed by users	5	per day
Number of active users	80	Million
Number of documents viewed in a day	<i>f</i> 400	Million
Number of documents viewed in a second	<i>f</i> 4630	per second
Bandwidth required for textual content per second	<i>f</i> 3.704	Gigabits per second(Gbps) ?
Bandwidth for image-based content per second	<i>f</i> 8.89	Gbps

Tt

⚙

nd		
Bandwidth for video content per second	^f 2.22	Gbps
Total outgoing bandwidth required	^f 14.81	Gbps

💡 See Detailed Calculations



Summarizing the bandwidth requirement

Note: The total bandwidth required is equal to the sum of incoming and outgoing traffic. $= 3 + 14.7 \approx 18\text{Gbps}$ approximately.

Number of servers estimation

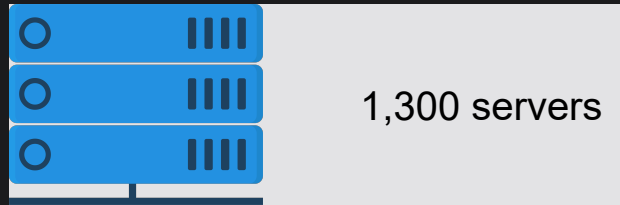
Since we have 80 million daily active users working on collaborative documents. Considering our assumption of using daily active users as a proxy for the number of requests per second to find the number of servers for peak load times, we get 80 million requests per second. Then, we use the following formula to calculate the number of servers:

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



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

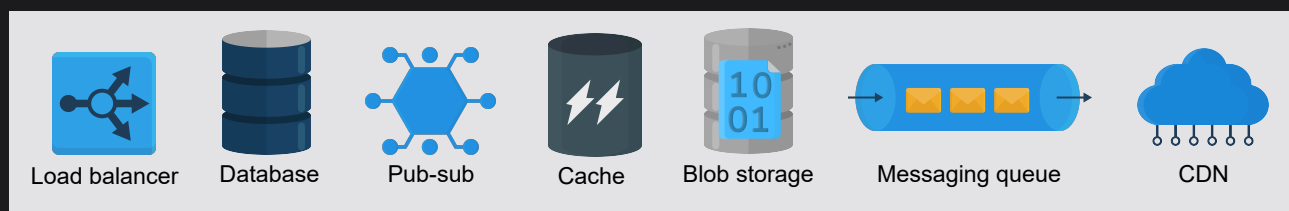
>
$$\text{Servers needed at peak load} = \frac{80 \text{ million}}{64,000} = 1250 \approx 1.3K \text{ servers}$$



The estimated number of servers required for Google Docs

Building blocks we will use

We'll use the following building blocks in designing the collaborative document editing service.



Building blocks required to be integrated in the design

- **Load balancers** will be the first point of contact for users.
- **Databases** will be needed to store several things including textual content, history of documents, user data, etc. For this purpose, we may need different types of databases.
- **Pub-sub** systems can complete tasks that can't be performed right away. We'll complete a number of tasks asynchronously in our design. Therefore, we'll use a pub-sub system.
- **Caching** will help us improve the performance of our design.
- **Blob storage** will store large files, such as images and videos.
- **A Queueing system** will queue editing operations requested by different users. Because many editing requests can't be performed simultaneously, we have to temporarily put them in a queue.



- A CDN can store frequently accessed media in a document. We can also put read-only documents that are frequently requested in a CDN.



← Back

System Design: Google Docs



Completed

Next →

Design of Google Docs

