Requirements of Yelp's Design

Learn about the requirements for a proximity service like Yelp.

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 identify the requirements of our system.

Functional requirements

The functional requirements of our systems are below:

- **User accounts**: Users will have accounts where they're able to perform different functionalities like log in, log out, add, delete, and update places' information.

> **Note:** There can be two types of users: business owners who can add their places on the platform and other users who can search,

view, and give a rating to a place.

- **Search**: The users should be able to search for nearby places or places of interest based on their GPS location (longitude, latitude) and/or the name of a place.
- **Feedback**: The users should be able to add a review about a place. The review can consist of images, text, and a rating.

Functional requirements

Non-functional requirements

The non-functional requirements of our systems are:

- **High availability**: The system should be highly available to the users.
- **Scalability**: The system should be able to scale up and down, depending on the number of requests. The number of requests can vary depending on the time and number of days. For example, there are usually more searches made at lunchtime than at midnight. Similarly, during tourist season, our system will receive more requests as compared to in other months of the year.
- **Consistency**: The system should be consistent for the users. All the users should have a consistent view of the data regarding places, reviews, and images.
- **Performance**: Upon searching, the system should respond with suggestions with minimal latency.

Resource estimation

Let's assume that we have:

- A total of 178 million unique users.
- 60 million daily active users.
- 500 million places.

Number of servers estimation

Considering our assumption (from Back-of-the-envelope Calculations) of using daily active users as a proxy for the number of requests per second, we get 60 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 at peak times are estimated as follows:

$$Servers\ needed\ at\ peak\ load = \frac{60\ million}{64,000} = 937.5 \approx 1K\ servers$$
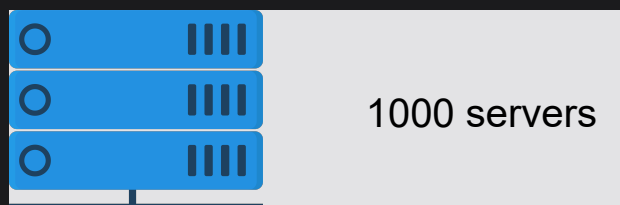
The number of servers required for Yelp

Storage estimation

Let's calculate the storage we need for our data. Let's make the following assumptions:

- We have a total of 500 million places.
- For each place, we need 1,296 Bytes of storage.
- We have one photo attached to each place, so we have 500 million photo
- For each photo, we need 280 Bytes of storage. Here, we consider the row size of the photo entity in the table, which contains a link to the actual photo in the blob store.
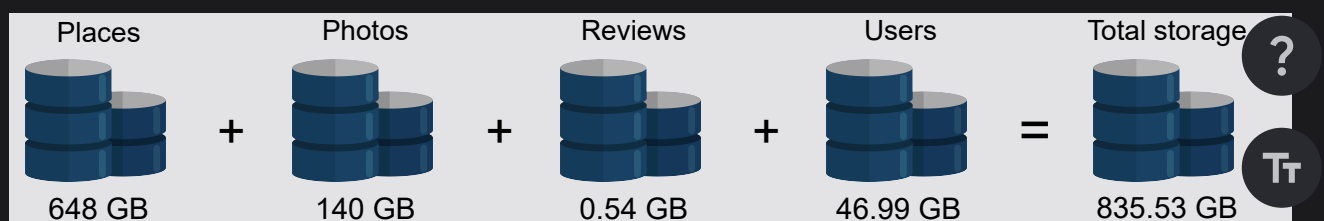
- At least 1 million reviews of different places are added daily.

- For each review, we need 537 Bytes of storage.

- We have a total of 178 million users.

- For each user, we need 264 Bytes of storage.

> **Note:** The Bytes used for each place, photo, review, and user are based on the database schema that we'll discuss in the next lesson.

The following calculater computes the total storage we need:

Estimating Storage Requirements

Type of information	Size Required by an Entity (in Bytes)	Count (in Millions)	Total Size (in GBs)
Place	1296	500	*f* 648
Photo	280	500	*f* 140
Review	537	1	*f* 0.54
User	264	178	*f* 46.99
Total Storage Required			*f* 835.53

Places		Photos		Reviews		Users		Total storage
648 GB	+	140 GB	+	0.54 GB	+	46.99 GB	=	835.53 GB

The total amount of storage required by Yelp

Bandwidth estimation

To estimate the bandwidth requirements for Yelp, we categorize the bandwidth calculation of incoming and outgoing traffic.

For incoming traffic, let's assume the following:

- On average, five places are added every day.
- For each place, we take up 1,296 Bytes.
- A photo of size 3 MB is also attached with each place. This is the size of the photo that we save in the blob store.
- One million reviews of different places are added every day.
- Each review, takes up 537 Bytes.

We divide the total size of information per day by 86,400 to convert it into per second bandwidth.

Estimating Incoming Bandwidth Requirements

Average Number of Places Added Daily	5
Storage Needed for Each Place (Bytes)	1296
Size of Photo (in MBs)	3
Total Size of Place Information (Bytes) *f*	15006480
Average Number of Reviews Added Daily (in Millions)	
Storage Needed for Each Review (Bytes)	537

Total Size of Reviews (Bytes) *f*	537000000
Total Incoming Bandwidth (KBps) *f*	6.39
Total Incoming Bandwidth (Kbps) *f*	51.12

For outgoing traffic, let's assume the following:

- A single search returns 20 places on average.
- Each place has a single photo attached to it that has an average size of 3 MB.
- Every returned entry contains the place and photo information.

Considering that there are 60 million active daily users, we come to the following estimations:

Estimating Outgoing Bandwidth Requirements

Average Number of Places Returned on Each Search Request	20
Size of Place (in Bytes)	1296

|	
Total Size of Place Information (Bytes) *f*	60025920
Outgoing Bandwidth Required for a Single Request (KBps) *f*	0.69
Outgoing Bandwidth Required for a Single Request (Kbps) *f*	5.52
Daily Active Users (in Millions)	60

Total Outgoing Bandwidth Required (Kbps)	*f*	331200000
Total Outgoing Bandwidth Required (Gbps)	*f*	331.2

We need a total of approximately $51\ Kbps$ of incoming traffic and approximately $331\ Gbps$ of outgoing, assuming that the uploaded content is not compressed.

Total bandwidth requirements = $51\ Kbps + 331\ Gbps \approx 331\ Gbps$.

Incoming bandwidth = 51 Kbps   +   Outgoing bandwidth = 331 Gbps   =   331 Gbps
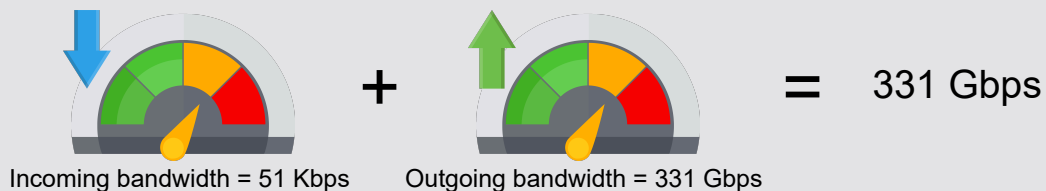
The total bandwidth required by Yelp

Building blocks we will use

The design process of Yelp utilizes many building blocks that have already been discussed in the initial chapters of the course. We'll consider the following concepts while designing Yelp:

Cache   Load balancer   Blob storage   Database
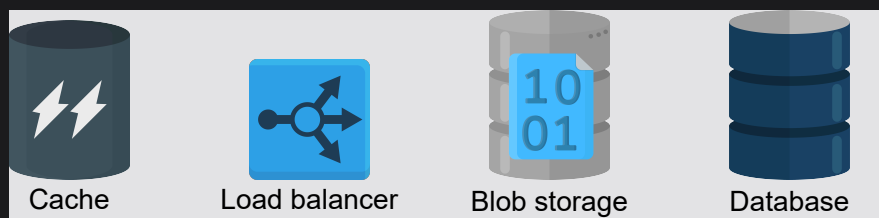
Building blocks in the high-level design of Yelp

- **Caching**: We'll use the cache to store information about popular places.
- **Load balancer**: We'll use the load balancer to manage the large amount of requests.
- **Blob storage**: We'll store images in the blob storage.

- **Database**: We'll store information about places and users in the database.

We'll also rely on **Google Maps** to understand the feature of searching for places within a particular radius.

Back

✔ Completed

Next →