Data given and assumptions on it
As we Have Data of around 5M movies

Considering 1 Movie ~= 1KB of data
Total space needed ~= 5GB scaling of data to 5X will needed around 25GB so storage is not a problem here

Now come to second Point which is 15M request every single day ~= 180 requests per second
Now consider in peek hours we might have 5X traffic ~= 900 requests per second

On this we should consider scaling of requests to 5X which goes to ~= 4500 requests per second in our peek hours

There are mainly two posible bottelnacks

1. Service level
2. DB level

Solution for Service level Bottleneck
We need to **horizontally scale** our service(s) and have to put **load balancer** before

DB level Bottleneck Solutions

1. We can assume that only 60% of our user go to 2nd page and like 40% will also go to 3rd page and after page 10 only 2% users visits
   So, for this we can use cache layer to serve fast and can prevent our persistent DB from large number of requests
2. As we know our system is read heavy
   - Movies will added least frequently then write
   - Only admin users can make Create,Update and Delete operations

   So, we can have Multiple Read Replicas of our database and our all read API will be served from read replicas

imdb

sunil saini | April 5, 2021