Content Delivery Networks

Background

In most large scale systems, there is a need to serve to a user a variety of static content - this includes HTML files, images, videos, or any other file type where the data is not expected to change frequently.

However, these files are typically large, and as a result serving them from our application servers can potentially be very slow, especially when the end user accessing them is far away. Additionally, there is no reason for a server to have to recompute these files every time, since they are static and should not be changing.

To solve these problems, companies use Content Delivery Networks (CDNs).

Storing Static Content

Where do we generally store static content? An object store like S3, Azure Blob storage, or Google Cloud Storage.

- Far more scalable than running a cluster dedicated to running a distributed file system such as HDFS
  - Unlike HDFS, files are stored using keys and values, as opposed to in a directory hierarchy
  - You only pay for the space that you use, as opposed to all the hardware in the cluster
  - Can scale storage independently from compute by nature of cloud services
  - Object accesses should be faster as flat structure allows avoiding use of a NameNode, no locks on data since writes are immutable (new writes to same key increases version number)
- However, there are also down sides to object stores
  - No support for multi-object transactions
  - If you want to run batch jobs/streaming, you need to first transfer data over the network to an HDFS cluster, do not receive data locality benefits

Object Stores Aren’t Good Enough

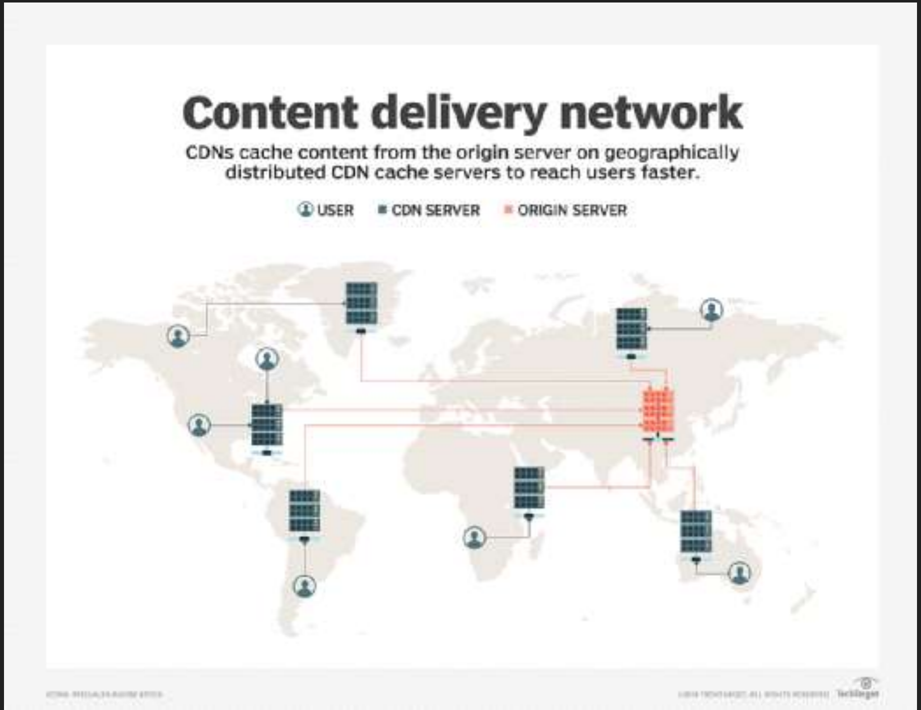
Even if you have your content in S3, these are generally large files and having to make network calls on each request is very inefficient. If the user is located far away from your application servers, the problem becomes even more of an issue.

Enter CDNs!

Content Delivery Networks

- A special type of geographically distributed cache that you should use to store static content from your servers
  - Partition the data such that data for specific geographic regions goes to the particular CDN that you want
  - CDNs can be either push or pull based
    - Push: Servers periodically send data to be stored in the CDN, and it serves all of the requests for that static content
      - Good for news sites, HTML home pages
    - Pull: Clients ask CDNs for content, and if they do not store it, get it from the server and then store it
      - Good for social media sites like YouTube, TikTok, Instagram

CDNs visualized



Content delivery network
CDNs cache content from the origin server on geographically distributed CDN cache servers to reach users faster.
USER
CDN SERVER
ORIGIN SERVER
TechTarget

CDNs Analyzed

Pros:

- Greatly reduces load on server for popular content
- Greatly increases speed to users around the world

Cons:

- May occasionally serve stale data, or slow down requests on cache misses
- Have to change static URLs in database to reflect that they are now stored by the CDN

Conclusion

In most large scale systems, there is a need to quickly serve static content such as HTML pages, images, and videos. While object storage systems such as AWS S3 are great for storing the original copies of these images, they are often times too slow to serve such large files to the end user.

Adding a content delivery network in conjunction with file storage not only takes a significant amount of load off of the application server, but it also greatly increases the speed of these requests for static content.

https://databricks.com/blog/2017/05/31/top-5-reasons-for-choosing-s3-over-hdfs.html

https://www.youtube.com/watch?v=farO15_0NUQ

https://www.youtube.com/watch?v=8zX0rue2Hic