Part 1
Brain Warm Up

Assume, we need to add the feature of **Download Zip** on our Citadel (Study Portal). Now, this feature will enable the users to download all the material regarding a particular course. Assume the data is stored in a directory tree structure, as it is shown on the website. Now, you have two options (or maybe more) to serve the zip file to the user, when he/she clicks on download zip.

(a) You generate a zip file of the complete folder of the course when, the user clicks Download Zip and then serve the file to the user.

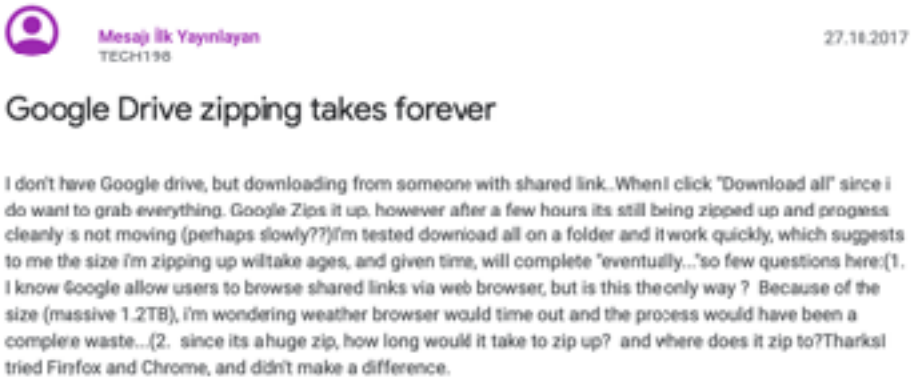
(b) You already saved the zip file of each course stored on the hard disk of the server, and then instantly serve the zip of the course.

(c) There may be some other better method...

Which of these methods do you think is more suitable for our particular need? Why do you think so? Is there another method, which will perform better heuristically?

Keep in mind, the computational load on the server to zip a say 2GB folder for a course, the hard disk limit, changing contents of a course when someone uploads a document and the basics of probability theory.

To practically find out the difference between the two methods, I started searching for a website that offers these methods.
On reading articles about zipping and compression online, I stumbled upon various forum and reddit posts about Google Drive taking forever to zip before starting a download.

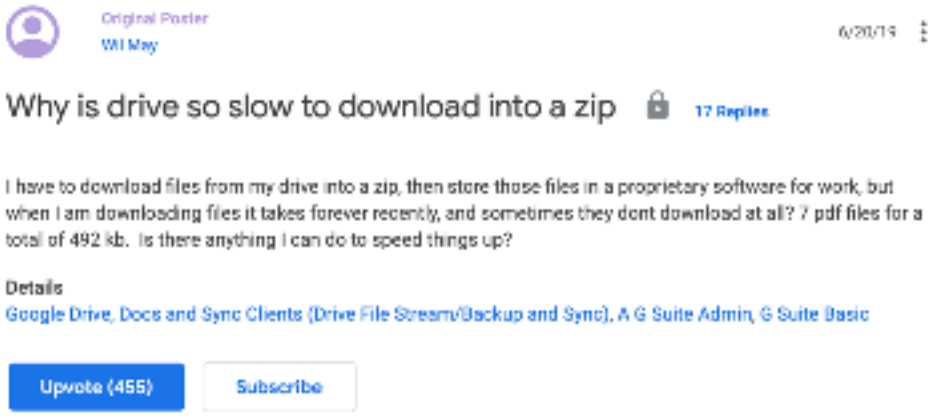


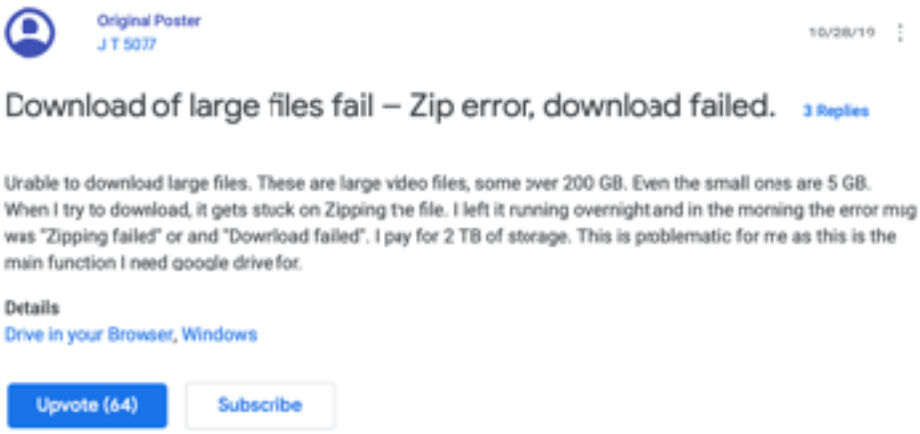
Mesajı İlk Yayınlayan
TECH198

27.10.2017

Google Drive zipping takes forever

I don't have Google drive, but downloading from someone with shared link..When I click "Download all" since i do want to grab everything. Google Zips it up, however after a few hours its still being zipped up and progress cleanly s not moving (perhaps slowly??)I'm tested download all on a folder and it work quickly, which suggests to me the size i'm zipping up willtake ages, and given time, will complete "eventually..."so few questions here:(1. I know Google allow users to browse shared links via web browser, but is this the only way ? Because of the size (massive 1.2TB), i'm wondering weather browser would time out and the process would have been a complete waste...(2. since its a huge zip, how long would it take to zip up? and where does it zip to?Thanks! tried Firefox and Chrome, and didn't make a difference.





Next, I started experimenting with Google Drive downloads.

1) Took around 50 files - pdfs, docx files, ppts, some jpegs and a video with a total size of 2.11GB saved on my drive.

On downloading all of them together, Google Drive by default makes them into one zip file, which on compression brought down the size to 2.08GB i.e. not a huge size difference was observed.

Zipping the folder took just over 6 mins and downloading the zipped folder took just over 6 mins as well.

Hence

a) Zipping on the server and downloading took roughly 13 mins.

While, b) Downloading the zipped folder directly took around 6 mins.

The internet speed varied from 8-10 Mbps throughout this time period.

2) Took files of cumulative size 728 MB

These took a total of 4 mins to zip and then download the zipped file.

Hence, time taken to zip+download increases roughly exponentially with time.
Which probably explains why terabytes of data took 'forever' to zip and download from Google Drive.

Next, I wanted to find out the difference in upload times for the 2 cases.

The 2.11GB folder was compressed to 2.08GB i.e. not a significant difference in size and took under a minute to compress locally.
(The default compression technique was used)

The upload times were virtually indistinguishable.

ANALYSIS and CONCLUSION

Let the server load on downloading a 2GB file approximately be *x* and on zipping it be *y, assuming x* and *y* are of the same order of magnitude.

There are around 68 courses available (from the BSW site as Citadel cannot be accessed outside campus).

Let the total memory required for all 68 courses be roughly 120GB.

Let around 20 people access the resources in an hour around the date of the exams.

Zipping and downloading would increase the load by about *20y* in an hour, a significant increase from just *20x*.

Hence, the better solution from above observations is to make available zipped files on the hard disk of the server, ready to download.
This however comes with the cons of a larger hard disk on the server, which would be pretty expensive for a task like this.
Another con would be the non availability of viewing single files to download from a course.
Another difficult for this method would stem from the regular uploading of new files for some courses, as updated zipped files would have to be uploaded every time a new file is added.

Also important is to take note of the resources at Google's disposal and the zip and download method still being annoying on Google Drive.

Zipping on the spot comes with a significant increase in operation time and especially so on lower speed networks.
However, it also comes in with an advantage as far as adding new files are concerned.

Best solution in this case would be a middle ground between the above 2 methods, based on the activity different course generate on the website.
The solution can be modeled based on a study of the usage of the portal over a period of time.
For example, First year courses contribute to a majority of traffic hence can be zipped and uploaded while other courses can have a zip and download available for them.