Project 2 – Assessing Hosting Options **Carsten Hood**
CSCE 489-599 Cloud Computing carstenhood@tamu.edu
Due Friday, October 20, 2017 UIN: 922009787

Report 1 – Assessing Application Hosting Options

1 Time Spent

I used all 8 hours allotted to this project distributed over three days from Friday 10/20 to Sunday 10/22.

I choose to apply 2 of my 4 free late-days to this assignment as I am submitting it 2 days after the due date.

2 Estimating Demands

Our first task prior to comparing hosting options is to extrapolate from the information provided by the client to form a conception of the client's application and to quantify hosting-related aspects of their business.

2.1 Estimating User Population

We know that the client predicts it will attract 5% of two universities' students in the first month and that it hopes to match Pinterest's adoption curve. First, we calculate 5% of the combined student populations of UT and TAMU to be 5870 users [1]. Including spillover into these students' circles as well as among our client's circles (e.g., among friends and family through social media, etc.), we estimate 7970 users in the first month. Then we examine Pinterest's user population history for reference; although no complete data could be found we were able to form a fairly accurate picture of Pinterest's first three years based on several online articles [2][3]:

Pinterest user population

month	*1 mo.*	*2 mo.*	*3 mo.*	*6 mo.*	*9 mo.*	*12 mo.*	*18 mo.*	*24 mo.*	*36 mo.*
date	Mar-10	Apr-10	May-10	Sep-10	Dec-10	Mar-11	Sep-11	Mar-12	Mar-13
users	1000	1500	2500	6000	10,000	150,000	2,023,000	19,000,000	50,000,000

Pinterest's user base began quite small but expanded roughly by a factor of 50% per month in its first year. We apply a similar model to the client's application. Note that in March 2011 Pinterest introduced an iPhone application, causing its user population to spike sharply [4]. Assuming our application will be available on all relevant major platforms early on, we anticipate no such sudden spike and apply a steadier and more modest growth model:

Estimated User Population

month	*1 mo.*	*2 mo.*	*3 mo.*	*6 mo.*	*9 mo.*	*12 mo.*	*18 mo.*	*24 mo.*	*36 mo.*
users	7,970	11,950	17,550	52,650	142,155	326,957	1,307,826	4,708,174	21,186,781

Our model predicts over 300,000 users in the first year, about 4.7 million in the second year, and 21.2 million at the end of the third year.

2.2 Estimating Data Uploaded

Classifying 10% of users as heavy content creators who upload 100 photos and 10 videos daily, 70% as average users who upload 5 photos and 0.3 videos daily, and 20% as pure consumers with no content provision, we apply our above user population model to calculate the following file upload numbers:

Estimated Files Uploaded

month	*1 mo.*	*2 mo.*	*3 mo.*	*6 mo.*	*9 mo.*	*12 mo.*	*18 mo.*	*24 mo.*	*36 mo.*
photos / day	107,595	161,325	236,925	710,775	1,919,093	4,413,913	17,655,651	63,560,344	286,021,546
videos / day	9,644	14,460	21,236	63,707	172,008	395,617	1,582,469	5,696,890	25,636,005

To estimate the digital size of these media files we consider their sources and purpose. Given the nature of our application, images are likely to be photographs as opposed to other types of images such as icons, memes, or other graphics. These photos will tend to be modern smartphone camera photos, about 1.9 MB each, but may also be higher quality photos or smaller compressed photos. They will be imported from Google Photos, iCloud, Facebook, Instagram, YouTube, and Pinterest accounts. Images are: (a) unlikely to be compressed in iCloud or Google Photos; (b) predictably 1080px by 1080px on Instagram; (c) slightly optimized to about 150 KB on average on Facebook; and (d) around 110 KB on average on Pinterest. Hence we assign an average photo size of 1.8 MB. Applying a similar process to videos, which we estimate to average 12 seconds in length (based on assuming a certain constitution of YouTube, Facebook, and Instagram videos and video clips stored with Google and iCloud), we assume an average of 6 MB per video. Multiplying the number of files by their average sizes we get:

Estimated Total Data Uploaded

month	*1 mo.*	*2 mo.*	*3 mo.*	*6 mo.*	*9 mo.*	*12 mo.*	*18 mo.*	*24 mo.*	*36 mo.*
TB per day	0.3	0.4	0.6	1.7	4.5	10.3	41.3	148.6	668.7
TB per month	7.5	11.3	16.6	49.8	134.6	309.6	1,238.2	4,457.7	20,059.6

2.3 Estimating Storage Demands

We can then factor in application-generated videos and approximate the amount of data that accumulates over the three-year interval for the purpose calculating storage expenses. We add a small multiplier to file sizes to account for metadata and algorithm-generated data. We assume 2% of all photos and 4% of videos are deleted by users each month. We don't assume a file or data limit per user, but note that some systematic means of deleting or incentivizing users to delete old photos may be desirable to lessen storage expenses.

Estimated Total Data Stored

month	*1 mo.*	*2 mo.*	*3 mo.*	*6 mo.*	*9 mo.*	*12 mo.*	*18 mo.*	*24 mo.*	*36 mo.*
total data (TB)	8	19	35	182	572	1,459	8,674	34,144	213,691

2.4 Estimating Data Output / Consumption

Considering the application and predicted distribution of user profiles, we guess users will view on average 9 videos daily. We assume an average application-generated video length of 30 seconds, similar to related videos on Facebook, Instagram, or Google Photos, suggesting an average video file size of 15 MB. Hence we have:

Estimated Data Consumed (Data Sent Out to Internet)

month	*1 mo.*	*2 mo.*	*3 mo.*	*6 mo.*	*9 mo.*	*12 mo.*	*18 mo.*	*24 mo.*	*36 mo.*
videos viewed / day	8.8E+04	1.3E+05	1.9E+05	5.8E+05	1.6E+06	3.6E+06	1.4E+07	5.2E+07	2.3E+08
video data output / day (TB)	1	2	3	9	23	54	216	777	3496
other data output / day (TB)	0	0	1	2	4	10	41	149	669
total output / month (TB)	47	70	103	310	838	1928	7712	27763	124934

3 Estimating In-House & Capital Expenses

Both the cloud-based and TAMU-based services are managed by separate parties with their own hardware and administrators. However, the client will still need specialized in-house personnel and tools for managing the backend software and interacting with the backend provider, either TAMU or AWS. We estimate that these costs will be similar between the cloud-based and TAMU-based options so we only calculate them once. We assume a monthly salary of $9,200 for in-house backend engineers and $9,500 for in-house backend managers [5], as well as certain capital and other costs per employee.

Estimated In-House Labor & Capital Expenses

month	*1 mo.*	*2 mo.*	*3 mo.*	*6 mo.*	*9 mo.*	*12 mo.*	*18 mo.*	*24 mo.*	*36 mo.*
num. in-house backend engineers	2	2	2	3	3	3	3	3	4
num. in-house backend managers	1	1	1	1	1	1	1	2	2
total monthly salaries	$27,900	$27,900	$27,900	$37,100	$37,100	$37,100	$37,100	$37,100	$46,300
monthly in-house expense cost	$3,690	$3,690	$3,690	$4,470	$4,470	$4,470	$4,470	$5,250	$6,030
monthly in-house capital cost	$2,040	$2,040	$2,040	$2,610	$2,610	$2,610	$2,610	$3,180	$3,750
total monthly in-house cost	$33,630	$33,630	$33,630	$44,180	$44,180	$44,180	$44,180	$45,530	$56,080

Note that this chart includes our sole capital costs, under *monthly in-house capital cost*, which would include in-house facilities, hardware, and other capital purchased and maintained by the company and subject to depreciation. All other costs (and hence all important costs) discussed are expense costs that do not result in assets to the client.

4 Cloud-Based Hosting

For our cloud-based option we consider using Amazon Web Services.

4.1 Cloud-Based Costs

We must consider the costs of storage, data transfer out, and processing. Data transfer into AWS servers from the Internet is typically free so we don't need to account for any cost associated with uploading data to the cloud [6].

To estimate AWS data storage costs we examine AWS's storage price rates and relationships with other large-scale firms. We guess that 85% of data will be accessed less frequently and can be stored at cheaper rates, resulting in the following monthly expense outlay:

Estimated Cloud Storage Expenses

month	*1 mo.*	*2 mo.*	*3 mo.*	*6 mo.*	*9 mo.*	*12 mo.*	*18 mo.*	*24 mo.*	*36 mo.*
standard rate ($/GB)	$0.0230	$0.0230	$0.0230	$0.0220	$0.0210	$0.0210	$0.0200	$0.0190	$0.0190
infreq. access rate ($/GB)	$0.0125	$0.0125	$0.0125	$0.0125	$0.0125	$0.0125	$0.0125	$0.0125	$0.0125
stand. storage cost / mo.	$26	$64	$120	$600	$1,802	$4,595	$26,022	$97,311	$609,018
infreq. storage cost / mo.	$80	$198	$370	$1,932	$6,080	$15,500	$92,161	$362,783	$2,270,462
total storage cost / mo.	$106	$263	$490	$2,532	$7,882	$20,095	$118,183	$460,095	$2,879,480

To estimate the cost of transferring data out of AWS, we apply Amazon's standard transfer and data retrieval rates up until their limit, which we are predicted to reach within our first year of growth [7]. For larger-scale data transfer later in our predicted growth model, we examine Netflix's custom arrangement with AWS and apply a similar rate structure [8]:

Estimated Cloud Data Transfer Out Expenses

month	*1 mo.*	*2 mo.*	*3 mo.*	*6 mo.*	*9 mo.*	*12 mo.*	*18 mo.*	*24 mo.*	*36 mo.*
stream rate ($/GB)	$0.085	$0.070	$0.070	$0.050	$0.040	$0.040	$0.035	$0.035	$0.030
streaming cost / mo.	$3,995	$4,933	$7,244	$15,523	$33,530	$77,120	$269,920	$971,711	$3,748,026
retrieval cost / mo.	$470	$705	$1,035	$3,105	$8,383	$19,280	$77,120	$277,632	$1,249,342

Since we are limited in our time to write this report, we estimate the significant monthly costs of the application's intensive processing on the cloud simply as a factor of the amount of data uploaded monthly, as determined by examining similar image processing software's costs in relation to amounts of image data [9]:

Estimated Cloud Processing Expenses

month	*1 mo.*	*2 mo.*	*3 mo.*	*6 mo.*	*9 mo.*	*12 mo.*	*18 mo.*	*24 mo.*	*36 mo.*
cost	$211,030	$316,412	$464,689	$1,394,067	$3,763,980	$8,657,154	$34,628,617	$124,663,021	$560,983,593

We sum these expenses, including expenses for backing up data in AWS and our above-estimated in-house costs, to achieve a total cost for hosting the client's backend on a cloud-based service:

Estimated Total Cloud-Based Hosting Expenses

month	*1 mo.*	*3 mo.*	*6 mo.*	*9 mo.*	*12 mo.*	*18 mo.*	*24 mo.*	*36 mo.*
storage	$106	$490	$2,532	$7,882	$20,095	$118,183	$460,095	$2,879,480
data retrieval	$470	$1,035	$3,105	$8,383	$19,280	$77,120	$277,632	$1,249,342
processing	$211,030	$464,689	$1,394,067	$3,763,980	$8,657,154	$34,628,617	$124,663,021	$560,983,593
back-ups	$377	$1,742	$9,091	$28,610	$72,940	$433,698	$1,707,216	$10,684,527
tot. AWS bill	$215,601	$473,458	$1,415,226	$3,813,775	$8,773,649	$35,093,839	$126,372,457	$568,860,441
in-house	$33,630	$33,630	$44,180	$44,180	$44,180	$44,180	$45,530	$56,080
total cloud	**$249,231**	**$507,088**	**$1,459,406**	**$3,857,955**	**$8,817,829**	**$35,138,019**	**$126,417,987**	**$568,916,521**

4.2 Cloud-Based Hosting Analysis

Hosting with AWS is reliable, predictable, flexible, and allows for dynamic and inexpensive scaling. Many massive-scale storage- and processing-intensive services similar to our client's application use AWS or similar services. Using AWS has few and far-fetched disadvantages: it means our client gains no capital assets for all its backend expenditures and risks unlikely but devastating harm should AWS fail or be compromised [10].

Our rough estimates attribute the greatest cloud hosting expense to our client's extremely resource-intensive proprietary media-processing algorithm. To remain profitable in face of this large expense, the client would need to optimize this process to a great degree. Overall, cloud-hosting, particularly using AWS, seems like a feasible and relatively limitless approach; only a very special offer could improve upon it.

5 TAMU-Based Hosting

The client's report describes a unique hosting offer by TAMU: free data transfer, free disaster recovery, and massively discounted processing and storage rates. From the client's brief we gather that TAMU is providing as many clusters as necessary and considers this offer "an investment", implying TAMU expects a share of ownership in return.

5.1 TAMU-Based Hosting expenses

Since we are low on report-writing time, we use our AWS calculations to facilitate guessing an industry rate for hosting, comprising utility, administrator, and data storage costs, and use our previous user population and data calculations to produce monthly cost estimates. We then apply TAMU's heavy 95% discounted rate to obtain an estimate for hosting on TAMU's powerful server clusters, and then apply further discounts to this rate to obtain estimates for hosting on TAMU's less powerful clusters when they are (a) twice as slow and (b) 33% slower than the first more powerful configuration. We then incorporate our above-determined in-house costs with the recommended more powerful TAMU rate to obtain a total cost for hosting the client's backend on TAMU's servers:

Estimated Total TAMU-Based Hosting Expenses

month	*1 mo.*	*3 mo.*	*6 mo.*	*9 mo.*	*12 mo.*	*18 mo.*	*24 mo.*	*36 mo.*
industry rate	$211,136	$465,179	$1,396,598	$3,771,862	$8,677,249	$34,746,799	$125,123,115	$563,863,073
***TAMU rate** (fast)*	$10,557	$23,259	$69,830	$188,593	$433,862	$1,737,340	$6,256,156	$28,193,154
(slower cluster)	$6,334	$13,955	$41,898	$113,156	$260,317	$1,042,404	$3,753,693	$16,915,892
(slowest cluster)	$7,390	$16,281	$48,881	$132,015	$303,704	$1,216,138	$4,379,309	$19,735,208
in-house costs	$33,630	$33,630	$44,180	$44,180	$44,180	$44,180	$45,530	$56,080
recom. total	**$44,187**	**$56,889**	**$114,010**	**$232,773**	**$478,042**	**$1,781,520**	**$6,301,686**	**$28,249,234**

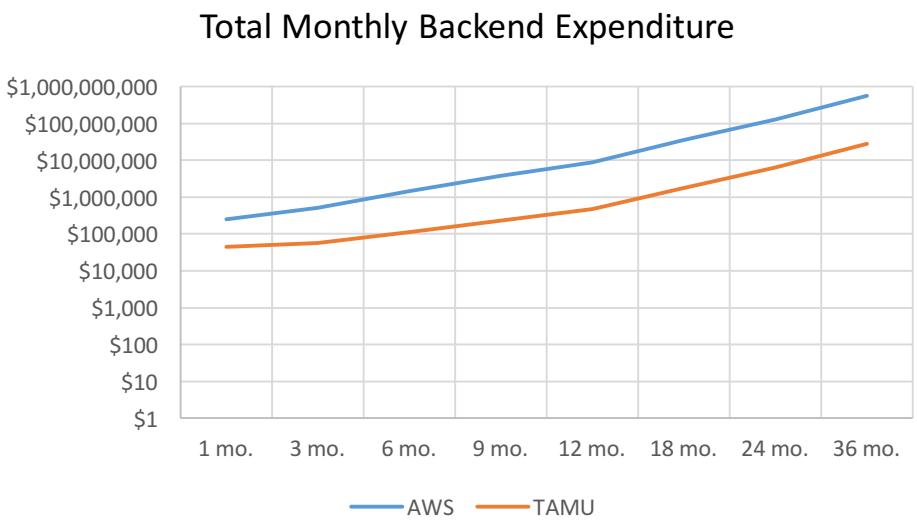
5.2 TAMU-Based Hosting Analysis

TAMU eliminates significant expenses by forgoing charging for network traffic and disaster recovery and provides a massively reduced rate for server usage. Cons of the TAMU-based approach include less reliability and flexibility than AWS and, more generally, future uncertainty for the period after the expiration of the 3-year deal. However, because both the client and TAMU would predictably benefit from maintaining a relationship, we see it as likely that a follow-up deal could be worked out prior to expiration of the first contract and do not see this disadvantage as prohibitive or even significant. In the worst case the client could transition or expand its backend to the cloud (as Netflix did) if it is unable to continue or expand its relationship with TAMU.

If using the TAMU hosting option, we recommend using and expanding upon TAMU's more powerful computer clusters. The client's application is extremely resource-intensive and would benefit from utilizing (and paying the price for) state-of-the-art server hardware.

6 Recommendation

We recommend using the TAMU-based hosting option due to its significant estimated savings. Based on our rough calculations, the TAMU-based approach appears drastically cheaper than the cloud-based approach. This is illustrated by comparing their final estimated monthly expense outlays:



That TAMU is said to be "investing" in the client implies that TAMU expects a share of ownership in our client's company in exchange for providing discounted hosting services. Therefore, even if the TAMU-based option provides significantly cheaper hosting, it may still be desirable for the client to use a cloud service like AWS to avoid paying the expense of forfeiting or diluting company ownership. The TAMU-based option appears significantly cheaper to the extent that we feel comfortable recommending accepting the TAMU offer and foregoing third-party cloud-based hosting for the present term; however, the client must make this decision based on the terms of any considered agreement with TAMU and its own leadership's financial and ownership considerations.

7 Reference Links

1. colleges.startclass.com/compare/4251-4254/Texas-A-And-M-University-College-Station-vs-The-University-of-Texas-at-Austin;
2. en.wikipedia.org/wiki/Pinterest#2009.E2.80.932011:_Founding;
3. techcrunch.com/2012/02/07/pinterest-monthly-uniques/;
4. businessinsider.com/inside-pinterest-an-overnight-success-four-years-in-the-making-2012-4?page=1;
5. glassdoor.com/Salaries/server-administrator-salary-SRCH_KO0,20.htm;
6. aws.amazon.com/s3/pricing/;
7. aws.amazon.com/ec2/pricing/on-demand/;
8. cloudyn.com/blog/estimate-your-cloud-data-transfer-costs-3-use-cases-netflix/;
9. calculator.s3.amazonaws.com/index.html;
10. npifinancial.com/blog/pros-and-cons-digging-into-amazon-web-services/;