

Capstone Cost Analysis

Amount of data: 80 mb/day = **2400 mb/month** = 28.8 gb/yr = **86.4 gb/3 yrs**

AWS Pricing Estimates:

AWS DynamoDB	Indexed Data Storage (gb/month)	Data Transfer In/Out (gb/month)	Free Tier?	Total Cost (\$/year)
1st year	2.4	2.4/1	yes	0
2nd year	2.4	2.4/1	yes	0
3rd year	2.4	2.4/1	yes	0
				0

DynamoDB free tier constraints (per month): 25 gb free storage, up to 200 million read/write requests, 2.5 million read requests from streams

DynamoDB is non-expiring on the free tier, meaning AWS customers can continue to use the services (under the limit) without charge.

<https://aws.amazon.com/dynamodb/pricing/>

AWS S3	Storage Pricing (gb/month)	Requests(get/put)/Data Transfer (Get #, Put #, Out gb)	Free Tier?	Total Cost (\$/year)
1st year	2.4	30,60,1	yes	0
2nd year	2.4(\$.023)	30(\$.004), 60(\$.005), 1(\$0)	no	0.032
3rd year	2.4(\$.023)	30(\$.004), 60(\$.005), 1(\$0)	no	0.032
				0.064

S3 free tier constraints (first year, usage each month): 5gb free storage, 20000 get requests, 2000 put requests, 15gb data transfer out

This cost estimates that get requests will be from S3 to spark (EMR) and the put requests will be from the python script to DynamoDB + the initial files sent to S3. However, I'm not sure if each request counts for more than one file. PUT is \$0.005/1000 requests, GET is \$0.004/10000 requests.

<https://aws.amazon.com/s3/pricing/>

AWS EMR (m4.large)	EC2 Price (\$/hr)	EMR Price (\$/hr)	Hrs/Month	Total Cost (\$/3 years)
1st year	0.12	0.03	30	54
2nd year	0.12	0.03	30	54
3rd year	0.12	0.03	30	54
				162

EMR runs on top of an EC2 cluster, so the price is the underlying EC2+EMR cost. Each day we need a cluster to load the data into the database, and to standardize the data with spark. This cost estimates that we can use the smallest cluster offered for a single hour a day, which should be plenty.

<https://aws.amazon.com/emr/pricing/>

Total AWS usage costs for 3 years: \$162.06

OSU VMWare Hosting Pricing Estimates:

VMWare Host	Storage Usage (per year)	Total Cost (\$/year)
1st year	28.8	$60 \times 12 = \$720$
2nd year	28.8	$(60 \times 4) + (90 \times 8) = 240 + 720 = \960
3rd year	28.8	$(90 \times 9) + (125 \times 3) = 810 + 375 = \1185
		2865

This cost estimates that we start with the lowest level VM and move up as we pass the storage tiers that they allow. Since we get 2.4 gb/month, we reach 28.8gb/year. That means that we can stay within the first tier VM (up to 40gb) for the first year. We pass the 40gb limit in the second year, and pass the 80gb limit of the second tier in the third year.

<http://is.oregonstate.edu/service/vmware-hosting>

Total VMWare host costs for 3 years: \$2865