

<https://github.com/ldkz2524/SampleAndrew.git> - Sample Git Repository Address

Cost Analysis on Amazon Glacier

A. Calculate a yearly cost of 1TB storage (cheapest)

- An archive is the smallest unit that Amazon Glacier has and it can store from a single byte up to 40TB. Each single archive will contain 32 kilobyte of metadata in addition to the data that is going to be stored.

The cheapest way to store 1TB for a year would be to simply upload 1TB into a single archive. Therefore the entire storage that an account uses for a year would be 1024.00003 gigabytes. The cheapest region in Amazon Glacier requires an account to pay 0.01 dollars per GB per month, so $1024.00003 * 0.01 * 12 = 122.88$ dollars per year.

For upload, the only additional cost that will be paid is if a data is deleted prior to 90 days. So as long as data is kept for longer than 90 days period, a user would only have to pay monthly storage cost.

1. Calculate the *cheapest* way to retrieve 1TB of 1TB (In a year)

- A user can only retrieve 5% of the data in Glacier for free every month. Therefore in the most ideal situation in this case, an account will be able to retrieve 60% of the data for free over span of a year.

To retrieve the data for the lowest rate possible, it is important to keep the peak retrieval rate as low as possible. In the most ideal case, the idea would be to balance the data that is going to be retrieved over a year equally. This translates to retrieving $1024\text{GB}/12\text{months} = 85.33\text{GB}$ per month.

Next the steps for calculating the retrieval fee are as follows:

- 1) Determine hourly retrieval rate for the data that you are retrieving (Data Size/Number of Hours)
- 2) Determine how much data can be retrieved in an hour for free (Daily Retrieval Allowance/ Number of Hours)
- 3) Subtract 1) by 2)
- 4) Multiply 3) by the retrieval fee per GB and by number of hours in a month (720 hours for a 30 days month)

With the assumption that a month is 30 days long:

- 1) $85.33\text{GB}/(24\text{ hours} * 30\text{ days}) = 0.1185185185$
- 2) $(1024\text{GB} * 0.05)/(24\text{ hours} * 30\text{ days}) = 0.0711111111$
- 3) $0.1185185185 - 0.0711111111 = 0.047407407$
- 4) $0.047407407 * 0.01 * 720\text{ hours} = 0.3413333328$
- 5) $0.3413 * 12 = 4.10$

The assumption made is that the region archive is stored in has the cheapest rate and the every month in a year is a 30 days month.

2. Calculate the most *expensive* way to retrieve 1TB of 1TB (In a year)

The most expensive way of retrieving the data would differ with different region, so it is essential to make some assumptions. The assumptions that will be made are that the archive will be stored in the same region as the previous answer and all months are 30 days each.

Since peak hourly retrieval rate is what determines the rate that is to be paid for each month, to make it the most expansive, the peak hourly rate just has to be high at one point. So the most expansive case will be to retrieve all data over 4 hour period on a single day, which is the minimum time unit for Glacier. The procedure to calculate would be similar to the previous example.

- 1) $1024\text{GB}/4 = 256\text{GB}/\text{hour}$
- 2) $(1024\text{GB} \cdot 0.05)/4 = 12.8\text{GB}/\text{hour}$
- 3) $256 - 12.8 = 243.2\text{GB}/\text{hour}$
- 4) $243.2\text{GB}/\text{hour} \cdot 0.01 \cdot 720 \text{ hours} = 1751.04 \text{ dollars.}$

So this method of retrieving 1TB over 4 hours would cost 1751.04 dollars.

B. Calculate a yearly cost of 10TB storage (cheapest)

With the same assumption made in the previous answers, the calculations are not going to be much difficult. Since the data size is 10TB, it can be stored in a single archive. Therefore total size of data to be stored in Glacier is 10240.00003GB.

$$10240.00003\text{GB} \cdot 0.01 \cdot 12 = 1228.8 \text{ dollars a year}$$

1. Calculate the *cheapest* way to retrieve 1TB of 10TB (In a year)

1TB out of 10TB can be retrieved from Glacier for free over two months, because 1TB is only 10% of 10TB data stored in Glacier and 5% can be retrieved each month. Therefore to be precise, 1TB can be retrieved for free over span of a year and even in span of two months.

2. Calculate the most *expensive* way to retrieve 1TB of 10TB (In a year)

The most expansive way of retrieving data would be to retrieve it within a 4 hour period of a single day. If the calculations are carried out,

- 1) $1024\text{GB}/4\text{hours} = 256\text{GB}/\text{hour}$
- 2) $(10240\text{GB} \cdot 0.05)/4 = 128\text{GB}/\text{hour}$
- 3) $256 - 128\text{GB}/\text{hour} = 128\text{GB}/\text{hour}$
- 4) $128\text{GB}/\text{hour} \cdot 0.01 \cdot 720 = 921.6 \text{ dollars}$

c. Draw some conclusion of best method for retrieval.

To conclude, the best method of retrieval would be to not go over the limit of 5% free retrieval allowance every month. If this is enough for the lab purposes, this would be the best solution possible and also if the retrieval is under 5% limit, the retrieval job doesn't have to be balanced over a month, so the data can be retrieved relatively fast.

However if there is a need to retrieve the data over the limit of 5%, it is very important for user to balance the retrieval job well over span of a month. Since Amazon Glacier bills a user every month, it is important to keep peak hourly rate low every month, which is the same as saying balancing retrieval job well over span of a month. The Amazon Glacier does not balance the retrieval job load for the user, so the users would have to program it themselves. As a final point, when it comes to billing, Amazon Glacier only considers the peak hourly retrieval rate, so it is important to keep the rate as low as possible. This also means that lab would have to plan ahead which data is going to be retrieved at certain point in time to keep this rate low.