Tier 2 Storage

Overview

This service does not support HIPAA or other protected data.

Research Technologies in partnership with UITS has implemented an AWS rental storage solution. This is necessitated by the expiration of free unlimited Google Drive storage. The documentation below will walk researchers through creating an S3 account which is managed by AWS Intelligent Tiering. After 90 days of nonuse, data will be moved to Glacier and after 90 additional days, will be moved to Deep Glacier. There will be no charge for data stored at either Glacier level, nor for any transfer charges. The data can be retrieved at any time, although it will take a while.

This AWS option is called Tier 2 which differs from Tier 1, the primary storage that is directly connected to the HPC clusters. Tier 1 is very fast, very expensive, and immediately available for active analyses. Tier 2 is intended for data not immediately undergoing active analyses and for backups (highly encouraged!). Researchers can use the software Globus to move data to Tier 2, and can also move data from other sources (called endpoints) like Google Drive. The data in Tier 2 will not be mounted on HPC, and so Globus will be used to move it back to Tier 1 if needed.

AWS storage is organized in buckets. One S3 intelligent tiering bucket is supported per KFS account. A PI could sponsor multiple buckets by submitting separate requests each with a unique KFS number, and then provide permissions as they see fit. Note this is different from Google Drive where anyone could create one.

For any support questions, our consultants use ServiceNow and can be reached with a support ticket.

Workflow

1. The PI will go to the Portal and request a special AWS allocation. They will need to provide KFS account information including the Department's financial contact for billing purposes.
2. Our infrastructure team will create the “S3 bucket”. Once the bucket is ready, the PI will be notified by email.
3. The PI and their group will set up a Globus endpoint by following the detailed instructions below.
4. Once the Globus endpoint is created, data can be moved between the new AWS account and Tier 1, Google Drive, or external data sources. General Globus usage information is here.
5. A bill will be generated monthly for S3 usage beyond the subsidized 1 TB.
Pricing

Overview

Part of this service is paid for by researchers and the rest is either subsidized or covered by UITS. The data stored in S3 will be billed monthly by AWS to the KFS account used when this is set up. Data in archival storage will be stored at no cost to the researcher. You will receive an email with detailed billing information when charges are made to your account.

Storage Costs

![Warning] Very small files (less than 128KB) are not subject to intelligent tiering and are not migrated to Glacier/Deep Glacier. This means they are permanently stored in the paid storage class. If you have many small files, we recommend making archives of your directories (.tar, .gz, .zip, etc) prior to uploading them to AWS. This will also reduce transfer times significantly.

<table>
<thead>
<tr>
<th>Tier</th>
<th>Cost to Researchers</th>
<th>Duration</th>
<th>Data Retrieval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard</td>
<td>$0 (First TB)</td>
<td>Three Months (if data not downloaded*)</td>
<td>Data can be retrieved immediately</td>
</tr>
<tr>
<td></td>
<td>$23/TB/Month** (data &gt; 1TB)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glacier</td>
<td>$0</td>
<td>Three months (if data not downloaded*)</td>
<td>A restore request must be submitted. Restores may take a few minutes to hours. Data may be transferred once restored.</td>
</tr>
<tr>
<td>Deep Glacier</td>
<td>$0</td>
<td>Unlimited (if data untouched*)</td>
<td>A restore request must be submitted. Restores may take a few hours to days. Data may be transferred once restored.</td>
</tr>
</tbody>
</table>

* Downloading files restores data to the Standard storage class and restarts the duration counter
** More up-to-date pricing information can be found on AWS’s website

Data Transfer Costs

Data movement costs are subsidized by UITS so researchers are not charged any AWS transfer fees.

Request Storage

Who can submit a request?

A group's PI is responsible for submitting a storage request unless they have an xdisk/storage delegate. Delegates may perform Tier 2 storage operations on behalf of their PI by clicking Switch Users and entering their PI's NetID in the user portal. PIs may add delegates by entering their group member's NetID in the user portal under Add Delegate.

First, log into the User Portal and navigate to the Storage tab at the top of the page. Select Submit Tier 2 Storage Request.
This will open a web form. Add your KFS number under **KFS Number** and the email address for the Department's financial contact under **Business contact email**. There will also be two optional fields: **Subaccount** and **Project**. These are used for tagging/reporting purposes in KFS billing. You can safely leave these entries blank if you're not sure what they are. Once you have completed the form, click **Send request**. The KFS number can be obtained from the same financial contact.

Submitting this form will open a ServiceNow ticket. Processing time may take up to a few days. Once your request has been completed, you will receive a confirmation email with a link to subscribe for account alerts (e.g., notifications for a sudden spike in usage).

[EXT]AWS Notification - Subscription Confirmation

```
C CW Alert topic for ua-t2-bucket-netid <no-reply@sns.amazonaws.com>
To: Lastname, Name - (netid)

You have chosen to subscribe to the topic:
arn:aws:sns:us-west-0:00000000000:S2-storage-netid-000000-SNSAlertTopic-ABCDEFG12345

To confirm this subscription, click or visit the link below (if this was in error no action is necessary):
Confirm subscription

Please do not reply directly to this email. If you wish to remove yourself from receiving all future SNS subscription confirmation requests please send an email to sns-opt-out
```
Checking Your Usage

AWS runs a batch update every night with the results being reported the following day. This means that if you have made any modifications to your allocation, your usage information will not be accurately reflected until the next batch update.

You may check your storage usage at any time in the User Portal. Navigate to the Storage tab, select View Tier 2 Storage, and click Query Usage.

<table>
<thead>
<tr>
<th>Bucket</th>
<th>Number of objects</th>
<th>Standard storage used (gigabytes)</th>
<th>Glacier storage used (gigabytes)</th>
<th>Deep archive storage used (gigabytes)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ua-rt-t2-netid</td>
<td>811</td>
<td>0.1316</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Close
Generate Access Keys

Access keys will allow you to connect your AWS bucket to the software Globus. This will enable you to make transfers directly between HPC and your Tier 2 storage allocation.

To generate an access key, log into the User Portal, navigate to the Storage tab, and select Regenerate IAM Access Key.

This will generate a KeyID and Secret Access Key used to establish the connection. Save these keys somewhere safe since once the window is closed, they cannot be retrieved. If you forget your keys, you can regenerate them.
Transferring Files

Globus

The easiest way to transfer files from AWS to HPC is using Globus. We have instructions in our Transferring Files page on how to set up an endpoint to access your AWS bucket as well as how to initiate file transfers.

Alternatives

Some other file transfer programs include rclone and Cyberduck.

Restoring Archived Data

Data that are not touched for at least 90 and 180 days are automatically retiered to archival storage (Glacier and Deep Glacier, respectively). Files stored in an archival state cannot be transferred out of AWS until they are restored. Restore requests can be submitted through the user portal under the Storage tab by clicking Restore Archived Tier 2 Storage Object:

This will open a box where you can enter the path to a file or directory in your bucket. Enter the path to the object you would like to restore:
Once you select an object, click **Send Request** to initiate the retrieval.

The time it takes for an object to be retrieved is dependent on its storage class. Objects in Glacier may take a few hours while objects in Deep Glacier may take up to a day or two. Once an object has been restored, it will move back up to the frequent access tier and can be downloaded using any transfer method you prefer.
FAQ Frequently Asked Questions

Official AWS FAQs are here: https://aws.amazon.com/s3/faqs/

A PI must submit their group’s storage request. The exception to this is if a group member has been designated xdisk/storage delegate. Delegates may submit a request on behalf of their PI by switching users in the user portal.

In most cases it will be in one business day.

Yes. Using Globus you can move data from Google Drive to S3.

You should check the Amazon site: https://aws.amazon.com/s3/pricing/?nc=sn&loc=4
As of March 2022:

- Frequent Access Tier, First 50 TB / Month $0.023 per GB
- Frequent Access Tier, Next 450 TB / Month $0.022 per GB
- Frequent Access Tier, Over 500 TB / Month $0.021 per GB

Not in the first release, but potentially as a future offering

Yes. The capacity used is recorded daily and the billing is a monthly average.

It is used for accounting purposes and used by your Department’s finance specialist.

Yes, you can use the CLI (command line interface) for information about your usage

No, but you can track the usage and remove any data that should not be there.

Glacier is effectively large, slow disks and Deep Glacier is tape storage.

Amazon S3 will likely support what you do. Perhaps our consultants can help to rework your workflows.

You will not be charged for data ingress, egress or other operations.

Yes

Yes, for any question we use ServiceNow and can be reached with a support ticket.

The max file size is 5TB based on: https://docs.aws.amazon.com/AutoScaling/latest/UserGuide/service Limits.html

This alert is sent to notify you whenever your storage usage grows by 10% relative to the previous week. This ensures that you’re aware of any unintended spikes and the potential resulting costs.