Overview

Research Technologies in partnership with UITS is implementing 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.

Pricing

Part of this service is paid for by researchers and the rest is either subsidized or covered by UITS. The data that is stored in S3 will be billed monthly by AWS to the KFS account used when this is set up. The first TB is free, meaning that it will be covered by UITS. The data that get migrated to Glacier or Deep Glacier is covered by UITS. Any transfer or other costs are covered by UITS. Refer to AWS’s website for more detailed, up-to-date information on storage costs.

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.

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).
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.

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.

Creating a Globus Collection

Log into https://www.globus.org/ using your university credentials. If you have never used Globus before, see our file transfer documentation to get started.

Under the Endpoints tab, enter UA AWS S3 in the search bar. In the results, you should see the name UA AWS S3 show up with the description Managed Public Endpoint. Click the endpoint's name to proceed.
Next, select the **Credentials** tab. If you are prompted for Authentication/Consent, click **Continue**

Then authenticate by selecting your Arizona email address, then **Allow**.

You will then be returned to the Credentials tab. From there, link to your AWS S3 Bucket by clicking the **Add Credential** button on the right.

In the form, enter the IAM Access Keys you generated in the user portal and click **Continue**
Next, go to the Collections tab and select Add a Guest Collection.

This will allow you to set up a connection to your personal AWS bucket and its contents.

First, where it says Authentication/Consent Required (as shown below on the left), select Continue. You will authenticate by selecting your university credentials and clicking Allow. This will return you to the Create New Guest Collection page.

Next, under Directory find your bucket by selecting Browse (as shown below on right). Once you've selected your bucket, give your collection a descriptive name under Display Name. Once you're done, click Create Collection at the bottom of the page to finish the process.
Once your Globus collection has been created, you can make transfers using the **File Manager**. For example, to transfer files between your AWS bucket and HPC, log into [https://www.globus.org/](https://www.globus.org/) and select File Manager tab from the left vertical bar. You should see a command-style window. If you don’t see a split screen, click the center option under Panels in the top right (shown in screenshot below). Click Search on one side to find your AWS bucket.

Transfer programs do not perform well when moving many, many files. Sometimes speeds may slow down by orders of magnitude and errors may occur. If you are trying to move a large directory, we recommend compressing it into a zip or tar archive for optimal results.

This will open a Collection Search tab. Select **Your Collections** and find the AWS collection you created.

Once you click your AWS collection, you will be brought back to the File Manager tab. Select the Search on the other side to open a connection to HPC.
This time, under **Collection Search**, search **UA HPC Filesystems** and click the result.

This will bring you back to the File Manager where you will be able to initiate transfers. To make a transfer, navigate through the filesystem until you find the desired file/directory, click it once to select it, and click **Start**. You may monitor the transfer's progress under **Activity** on the left-hand side.
A PI must submit their group's storage request. The exception to this is if a group member has been designated "disk/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/AmazonS3/latest/userguide/qfacts.html

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