11falsenonelisttrue

storageN

The use of `storageN` within these documents indicates that any storage platform can be used.

Current available storage platforms:

storage1

storage2

What is this Documentation?

This documentation will cover doing file transfers with gsutil over our dedicated fiber interconnect in order to download data from sources that use Google storage.

Quick Start

1. Login to the Compute platform

2. Set up Google Account variable

Account Information:

This is the account that has been granted access to the data by the data owner.

This is not necessarily just an email address.

3. Login with gcloud

Follow the URL that you are given.

It will ask your to sign into Google Cloud SDK. Use the email granted permission by the data owner here, e.g. washukey@wustl.edu.

This will take you to a WashU authentication page if it's your WashU email. Put in your information for the WashU sign in you normally would.

It will then ask for access to your account, click Allow.

It will give you a code, you will need to paste this code back into the terminal you are working in.

You can confirm this worked with the following command:

Multiple Accounts:

If this account is the only account listed, it will by default be the "active" one.

If there are multiple accounts, you can use the following to set the one being used active.

4. Transferring the Data

Set the following variables needed for the transfer.

//path/from/data/provider:

//path/from/data/provider is the location of the data on Google Storage, as provided by the group sharing the data.

You will need to set the following variables

You need to launch a bsub job to use google cloud tools

If you are a member of more than one compute group, you will be prompted to specify an LSF User Group with -G group_name or by setting the LSB_SUB_USER_GROUP variable.

The following command will run a trial of the transfer to make sure it works.

Once the test is complete and there are no issues, run the transfer by removing the -n option.

If there are problems or the transfer locks up, you can safely restart the transfer without losing progress as it will continue from where it was stopped just like normal rsync.

If you are transferring a large number of small files, using parallel transfers may work better.

The following command will run a transfer in parallel.

Using the parallel option is done so with the knowledge that it can be error prone.