Aspera Transfer Guide

National Center for Biotechnology Information (NCBI)
National Library of Medicine
Version 2.7, 2 February 2010

Contents

Notice

Reference herein to any specific commercial products, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government, and shall not be used for advertising or product endorsement purposes.

1 Overview

This application note provides instructions on the use and installation of Aspera Connect for high throughput file transfer with NCBI. There are now many cases where large file transfers, greater than 1 gigabyte (Gb), are commonplace and a single download session may involve hundreds of such files. As the sizes of the datasets have increased, we have found that the traditional methods of *ftp* or *http* do not have the performance characteristics needed to support this load of data.

Requirements for large scale data transfer over the internet include high bandwidth, auto checksum, recursive copy, and security based on strong keys. NCBI has chosen to use a product from Aspera, Inc because of these improved data transfer characteristics. FTP and HTTP access will continue to be available to retrieve files. Instructions are provided below for investigators who wish to use this newer, high-performance data transfer technology.

NCBI also is open to using additional products with the appropriate performance characteristics.

1.1 Scope

This document is intended for users transferring large data files from NCBI.

1.2 Revision History

2.1 Draft A: 11 May 2009	Modified for general NCBI use (donp)
2.2 Draft B: 13 May 2009	Comments from Dima and Janet (donp)
2.5 8 Oct 2009	Modified bulk download example (sandrobe)
2.6 7 Dec 2009	Referenced Aspera v2.3.1 as current version

2 Aspera

2.1 The fasp Protocol

The FASP protocol from Aspera (www.asperasoft.com) uses UDP, eliminating the latency issues seen with TCP, and provides bandwidth up to 1 gigabit per second (Gbps) to transfer data. It has a restart capability if data transfer is interrupted midstream and is well behaved, so if there is other data traffic on your network connections, it will back off in order to avoid starving other protocols. We have seen effective throughput up to 800 megabits per second (Mbps) to a single site.

The fasp protocol uses UDP port 33001-33009 for data transfer and you may need to contact your IT security staff if this port is not open to NCBI through your institutional firewalls.

NCBI is implementing Aspera for two use cases, occasional users who download files for direct use (Aspera Connect), and bulk users who will be downloading large amounts of data (ascp).
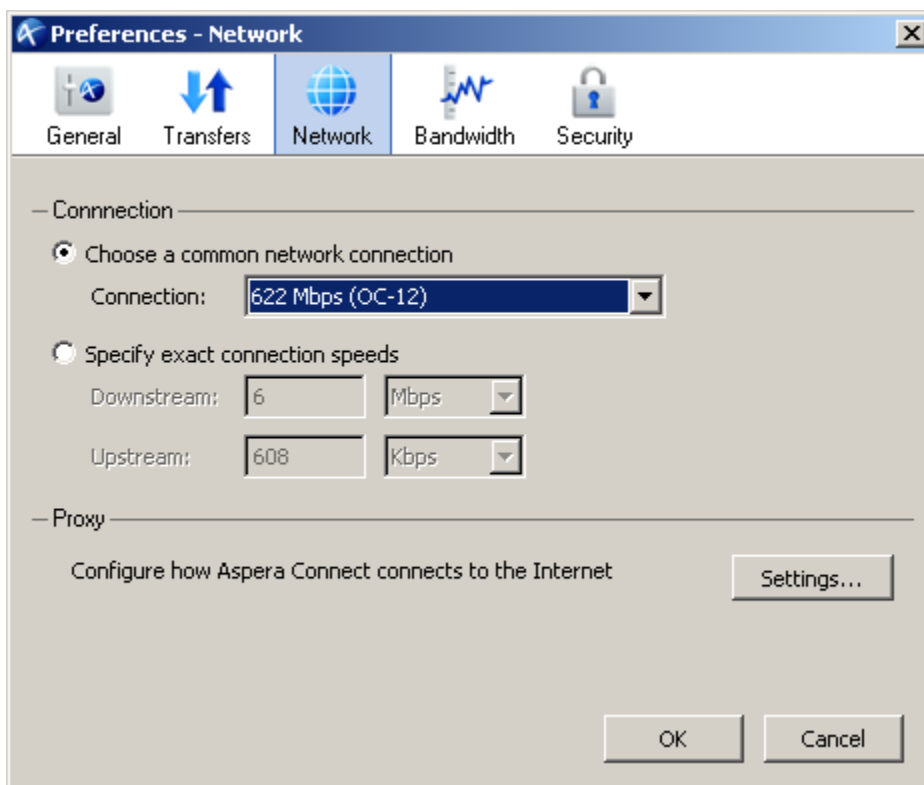
2.2 Aspera Connect

Aspera Connect is software that allows download and upload via a web plugin for popular browsers on machines running Linux, Windows, and Macintosh. The software also includes a command line tool that allows scripted data transfer. The software client is free for NCBI site users for the purpose of exchanging data with NCBI.

Download and install Aspera Connect software from:
http://www.asperasoft.com/downloads

Select the Connect product for your browser and platform. With version 2.3.1, disk throttling is enabled by default. Please upgrade to the current version of Aspera Connect to ensure correct performance.

2.2.1 Aspera Connect Configuration

To change, right click the Aspera icon from the system tray. Select 'networks' and update the connection speed (e.g. 622Mbps) to your network bandwidth. As an initial start, try 45Mbps and then go higher. You may find that you are limited by the performance of your local storage, so a few trials may be necessary to find the optimal rate. A standard desktop tends to top out around 250 Mbps.

```
Preferences - Network                                          [X]

  [General]  [Transfers]  [Network]  [Bandwidth]  [Security]

  — Connnection —————————————————————————————————
    (•) Choose a common network connection
         Connection:  [622 Mbps (OC-12)            ▼]

    ( ) Specify exact connection speeds
         Downstream:  [6    ]  [Mbps ▼]
         Upstream:    [608  ]  [Kbps ▼]

  — Proxy ——————————————————————————————————————
    Configure how Aspera Connect connects to the Internet   [Settings...]


                                              [  OK  ]  [ Cancel ]
```

2.3 Downloading Data with Aspera Connect (from browser)

Once the plugin has been installed in your browser, you may download file(s)/directory(ies) from NCBI using Aspera. Example: In your browser window, go to
http://www.ncbi.nlm.nih.gov/projects/faspftp/1000genomes/
Select a file (CHANGELOG is a good file to test with, with a left mouse click).
Click "Save" to begin saving the data. You will be prompted to select where the file is to be saved. For example:
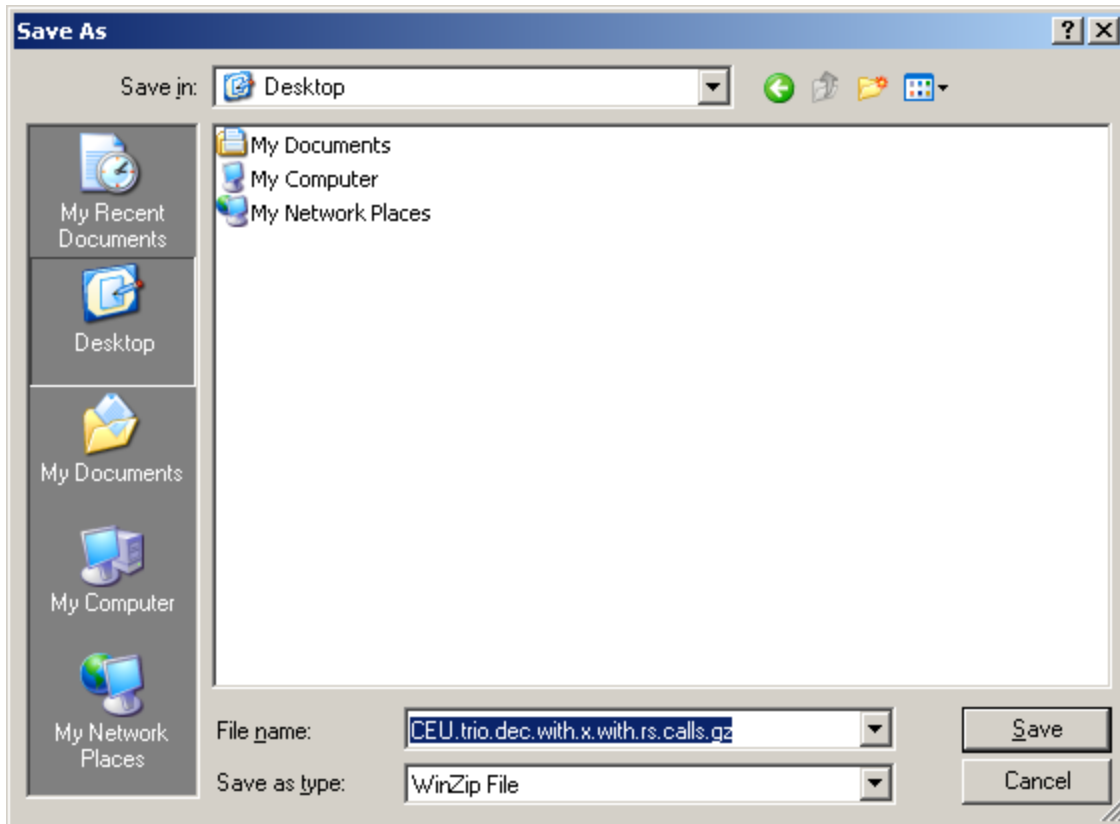
You can download full directories or a single file at a time.

2.4 Using ascp for Bulk Transfers

The command line program *ascp* is a utility delivered along with the AsperaConnect product.

- You can run the *ascp* program with the following parameter settings:
    - −Q (for adaptive flow control)
    - −l (maximum bandwidth of request, try 200M and go up from there)
    - −r recursive copy
    - −i <private key file>

- Try experimental transfers starting at 200 Mbps and working up to 400-500 Mbps. Select the bandwidth setting that gives good performance with unattended operation.
- `ascp -i <private key file>  -Q —l100m   `anonftp@ftp-private.ncbi.nlm.nih.gov:/ <files` to transfer> <local destination>`
where
<private key file> ::= fully qualified path & file name where the generated private key was saved.

<files(s) to transfer>   = names of files to transfer (including path)
<local destination path>
100M   = tunable mbit/sec bandwidth

- The *ascp* command on Microsoft Windows is located by default in C:\Program Files\Aspera\Aspera Connect\bin\ascp
- The *ascp* program on Mac in located at /Applications/Aspera Connect.app/Contents/Resources/ascp The *ascp* program on Linux is located at /opt/aspera/bin/ascp
- It is possible to run *ascp* in an autonomous, unattended manner that does not require repeated login.The asperaweb_id_dsa.*putt*y is in the C:\Program Files\Aspera\Aspera Connect \etc directory on Windows platform and is in /Applications/Aspera Connect.app/Contents/Resources/ on Mac OS X.

Additional information is available at the Aspera Web site: Windows, Mac OS X

For additional assistance, please contact the NCBI Help desk at info@ncbi.nlm.nih.gov