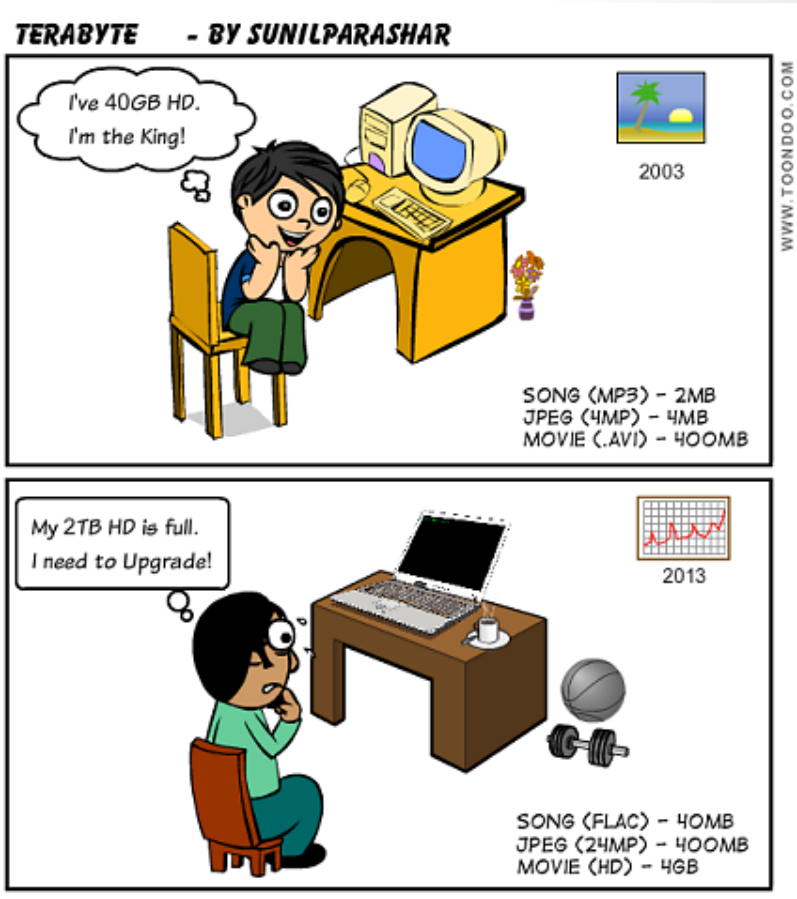
Data Transfer

Shelley Knuth
shelley.knuth@colorado.edu

www.rc.colorado.edu

Outline

- The meaning of data transfer
- How much is too much data?
- Data sizes
- Various Transfer methods
- Globus



Transferring Data

- This talk will outline the various and most common methods of transferring data between locations
- What is data?
  - Anything!
  - Text documents, Excel spreadsheets, Powerpoint files, binary files, images…anything that is being moved
- What is important for data transfer?
  - Speed
  - Security
  - Integrity

Transferring Data

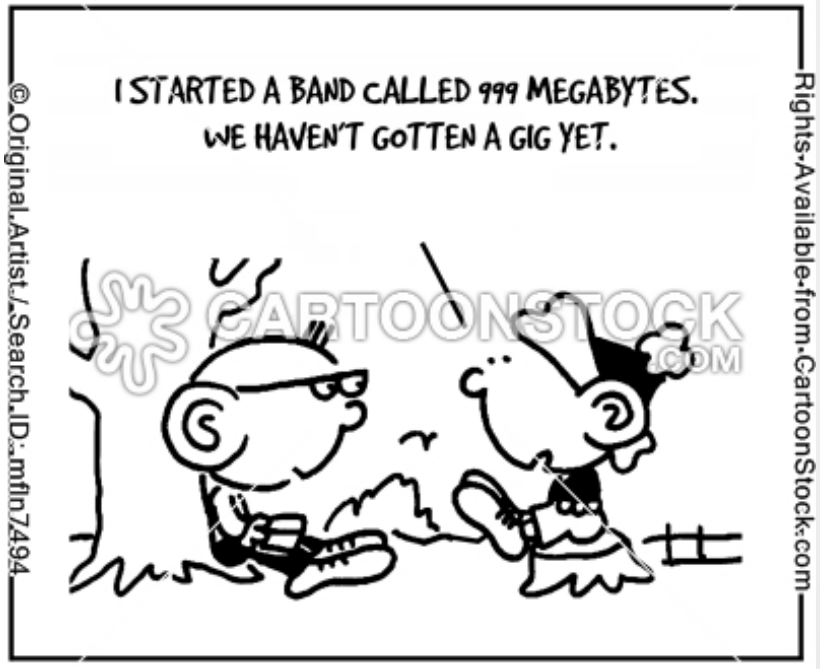
- Here we are discussing the movement of a file or files from one location to another
  - Generally one computer to another computer at a remote location
  - Either on the same network or separate
- The files are converted to packets then transferred over the network

XKCD

Data Sizes

- What's a GB? TB?

- 1 byte: 8 bits, the size of one alphanumeric character
- 1 KB: 1024 bytes
- 1 MB: 1024 KB
- 1 GB: 1024 MB
- 1 TB: 1024 GB
- 1 PB: 1024 TB



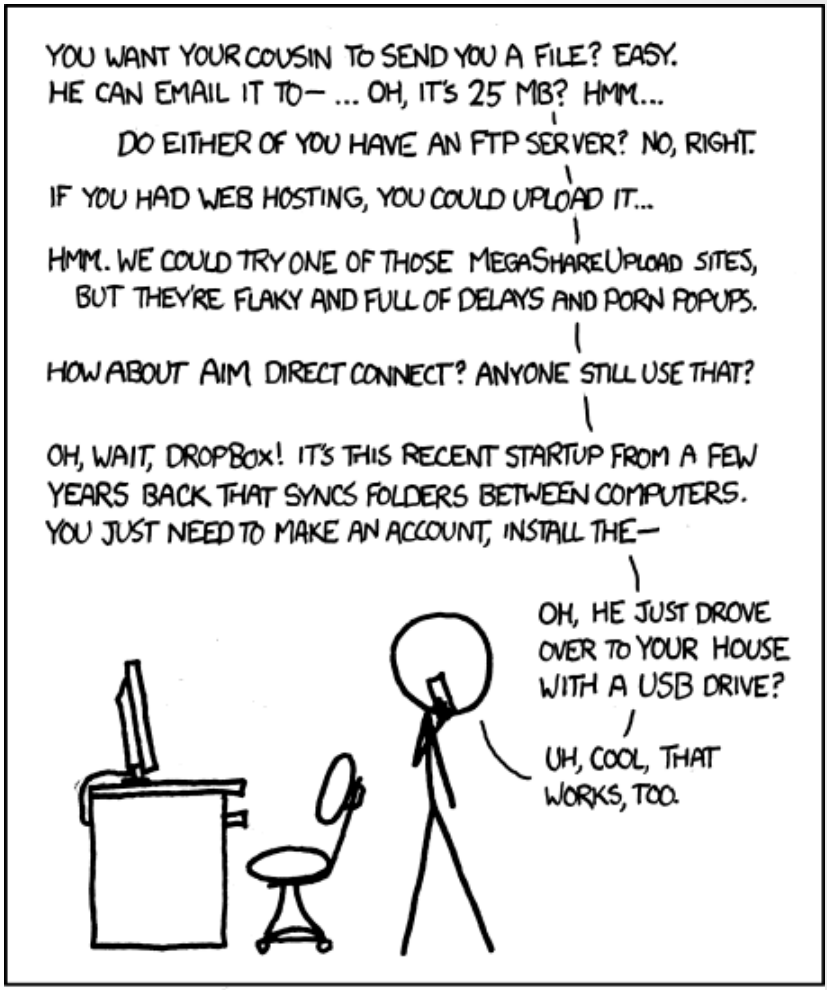
Data Sizes - Representation

- The average size of a single keystroke is 1 byte (8 bits)
- A line of text is 70 bytes
- 5 page word document is ~30 KB
- A typical iPhone photo: 500 KB
- 1 minute of CD quality audio: 1 MB
- Typical maximum email size: 10 MB
- MODIS Level 1B satellite file: 700 MB

http://www.gn.apc.org/support/understanding-file-sizes

Typical Data Transfer Rates

- Morse code: 0.05 Kbit/s
- Dial up modem: 40 Kbit/s
- DSL: 40 Mbit/s
- Cable modem: 100 Mbit/s
- Ethernet: 1 Gbit/s
- Wifi: 60 Mbit/s
- 4G Cell phone: 12 Mbit/s



I LIKE HOW WE'VE HAD THE INTERNET FOR DECADES, YET "SENDING FILES" IS SOMETHING EARLY ADOPTERS ARE STILL FIGURING OUT HOW TO DO.

XKCD

Various Transfer Methods

- There have been many means for transferring data over the years:
  - rcp
  - scp
  - ftp
  - wget
  - sftp
  - Bit torrent
  - rsync
- Some are in still in use and some are past their lifetime
- Let's cover some of these protocols in more detail

scp

- Secure copy
- Transfers files using encryption
- Example code:

scp file.txt username@remote_host.com:/home/username

Or, in reverse:

scp username@remote_host.com:/home/username/file.txt /home/username/

FTP

- File Transfer Protocol
- Transfer files, plus:
  - List files in directories
  - Resume uninterrupted transfers
  - Remote file manipulation
- Web browser:

  ftp://amrc.ssec.wisc.edu (Univ. WI Antarctic atmospheric data)
- Command line:

  ```
  ftp amrc.ssec.wisc.edu
  ```
- FileZilla

SFTP

- Secure File Transfer Protocol
- Like scp uses ssh to transfer files
- More secure and encrypted
- Becoming the standard
- Difference from scp:
  - Scp only allows transfer of files
  - FTP can list files, delete, etc

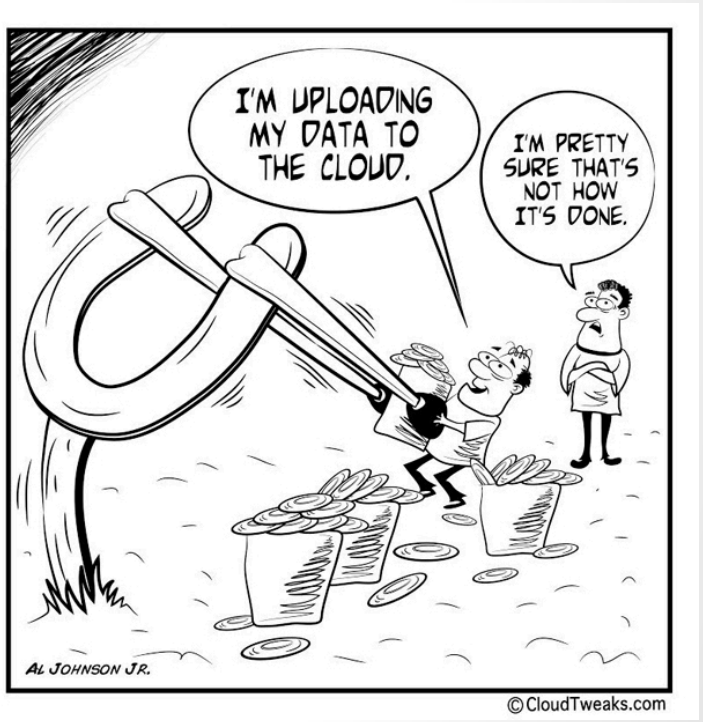
Rsync

- Rsync is a way to copy and synchronize two computers
- Faster than scp because it downloads only the differences between files (after initial download)
- Can use encryption
- Uses compression to move files
- Tons of options
- If lost connection, will resume where left off
  - Scp will not

Rsync options source destination

“The Cloud”

- What is this ambiguous cloud?
  - It’s a network of servers that run applications, deliver services, or store data
  - It is a good way to keep costs down and have data available wherever, whenever
  - Security issues?
  - Who owns your data?
  - Cost savings
    - Hardware, buildings, IT staff



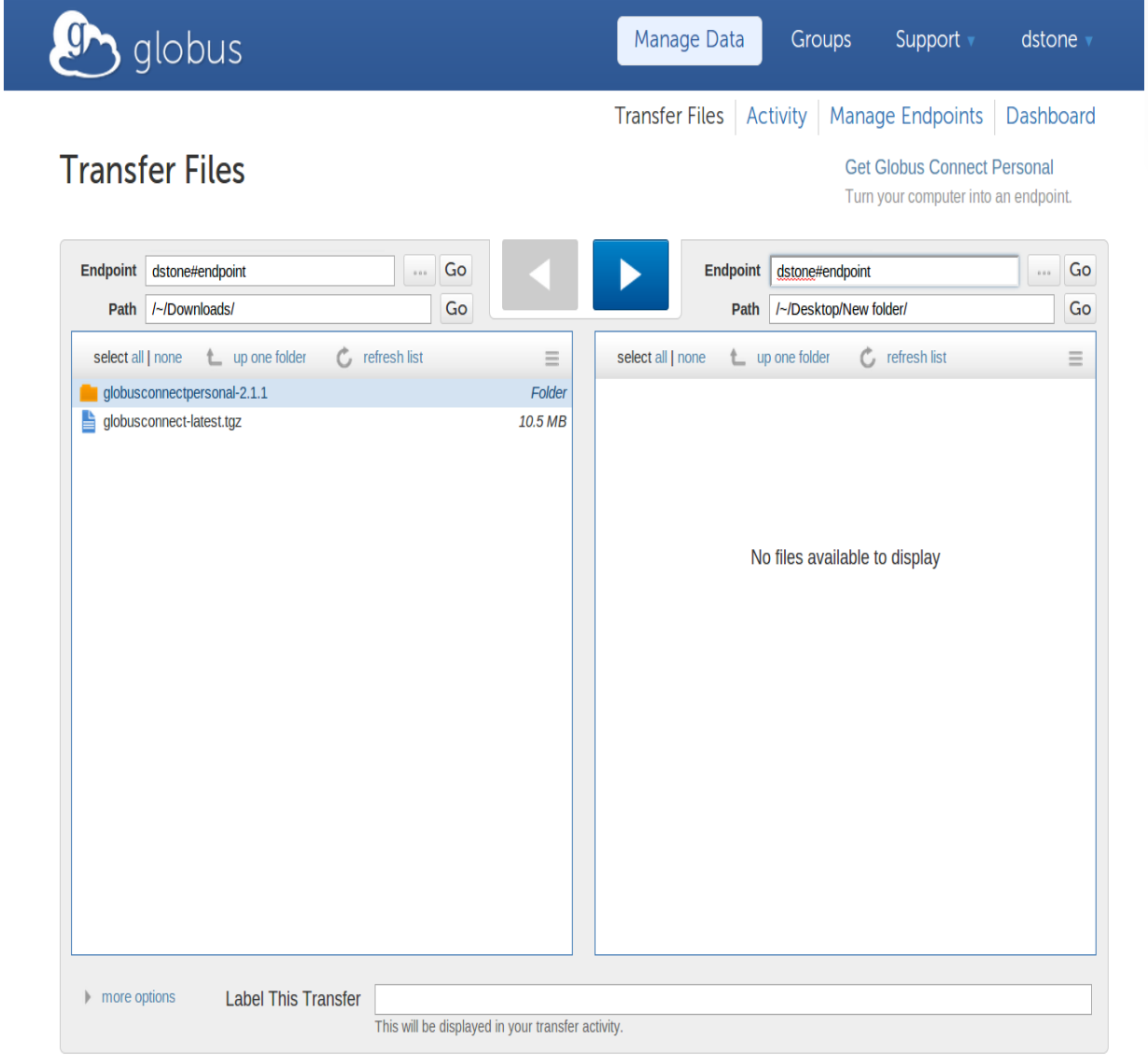
Globus

- Globus is our preferred method of data transfer
- Designed with researchers in mind
- End points between computers make for easy data transfer with an easy to use interface
  - Endpoints are different locations that data can be moved to/from
  - Personal or multi-user
- Scripting in use also if don't want to use GUI

www.globus.org

Globus

- Preserves the integrity of data
  - Compares checksums
  - Resumes data transfer if interrupted
- Fast transfer of large data sets
- Globus can be set up to easily share data among collaborators



You should now be able to find the endpoint you just created and start transferring files right to your computer

References Used:

- http://www.tecmint.com/rsync-local-remote-file-synchronization-commands/
- http://www.admin-magazine.com/HPC/Articles/Moving-Your-Data-It-s-Not-Always-Pleasant
- http://www.gn.apc.org/support/understanding-file-sizes
- compnetworking.about.com
- http://dem.nv.gov/uploadedFiles/demnvgov/content/NCSC/LTE-BroadbandAndPublicSafetyPrimer.pdf