Data transfer, storage and security considerations
Background

Example: Havasupai

- Native American tribe in US
  - High incidence of Type 2 diabetes
  - Approached ASU

- Original study done

- Later, more detailed genotyping
  - Standard analyses reported (amongst other results)
    - Inbreeding coefficient
    - Population bottleneck

- Havasupai sued ASU and were awarded damages
<table>
<thead>
<tr>
<th>startDate</th>
<th>End date</th>
<th>Source</th>
<th>Destination</th>
<th>Status</th>
<th>Transferred</th>
<th>Speed</th>
<th>from Email</th>
</tr>
</thead>
<tbody>
<tr>
<td>2016-01-29 09:25:04Z</td>
<td>2016-01-29 09:25:29Z</td>
<td>ppanchal#hgs-c-xfer</td>
<td>ncsa#BlueWaters</td>
<td>FAILED</td>
<td>0.0B</td>
<td>0.0B/s</td>
<td>Globus Notification <a href="mailto:no-reply@globus.org">no-reply@globus.org</a></td>
</tr>
<tr>
<td>2016-01-29 08:49:23Z</td>
<td>2016-02-02 19:30:12Z</td>
<td>ncsa#BlueWaters</td>
<td>grbot#H3ABioNetCentral</td>
<td>SUCCEEDED</td>
<td>3.8TB</td>
<td>84Mb/s</td>
<td>Globus Notification <a href="mailto:no-reply@globus.org">no-reply@globus.org</a></td>
</tr>
<tr>
<td>2016-01-27 09:04:39Z</td>
<td>2016-01-27 09:04:49Z</td>
<td>ncsa#BlueWaters</td>
<td>grbot#nautlius2</td>
<td>SUCCEEDED</td>
<td>12kB</td>
<td>9.4kB/s</td>
<td>Globus Notification <a href="mailto:no-reply@globus.org">no-reply@globus.org</a></td>
</tr>
<tr>
<td>2016-01-27 09:04:12Z</td>
<td>2016-01-27 09:04:24Z</td>
<td>ncsa#BlueWaters</td>
<td>grbot#nautlius2</td>
<td>SUCCEEDED</td>
<td>8.0kB</td>
<td>6.3kB/s</td>
<td>Globus Notification <a href="mailto:no-reply@globus.org">no-reply@globus.org</a></td>
</tr>
<tr>
<td>2016-01-27 09:04:00Z</td>
<td>2016-01-27 09:04:13Z</td>
<td>ncsa#BlueWaters</td>
<td>grbot#nautlius2</td>
<td>SUCCEEDED</td>
<td>11kB</td>
<td>7.3kB/s</td>
<td>Globus Notification <a href="mailto:no-reply@globus.org">no-reply@globus.org</a></td>
</tr>
<tr>
<td>2016-01-27 09:03:40Z</td>
<td>2016-01-27 09:03:53Z</td>
<td>ncsa#BlueWaters</td>
<td>grbot#nautlius2</td>
<td>SUCCEEDED</td>
<td>13kB</td>
<td>9.4kB/s</td>
<td>Globus Notification <a href="mailto:no-reply@globus.org">no-reply@globus.org</a></td>
</tr>
<tr>
<td>2016-01-27 09:03:22Z</td>
<td>2016-01-27 09:03:37Z</td>
<td>ncsa#BlueWaters</td>
<td>grbot#nautlius2</td>
<td>SUCCEEDED</td>
<td>7.5kB</td>
<td>4.2kB/s</td>
<td>Globus Notification <a href="mailto:no-reply@globus.org">no-reply@globus.org</a></td>
</tr>
<tr>
<td>2016-01-25 08:58:13Z</td>
<td>2016-01-25 09:01:15Z</td>
<td>ncsa#BlueWaters</td>
<td>grbot#H3ABioNetCentral</td>
<td>SUCCEEDED</td>
<td>52kB</td>
<td>2.1kB/s</td>
<td>Globus Notification <a href="mailto:no-reply@globus.org">no-reply@globus.org</a></td>
</tr>
<tr>
<td>2016-01-25 08:57:42Z</td>
<td>2016-01-25 09:01:40Z</td>
<td>ncsa#BlueWaters</td>
<td>grbot#H3ABioNetCentral</td>
<td>SUCCEEDED</td>
<td>41kB</td>
<td>1.0kB/s</td>
<td>Globus Notification <a href="mailto:no-reply@globus.org">no-reply@globus.org</a></td>
</tr>
<tr>
<td>2016-01-25 08:56:50Z</td>
<td>2016-01-25 09:02:22Z</td>
<td>ncsa#BlueWaters</td>
<td>grbot#H3ABioNetCentral</td>
<td>SUCCEEDED</td>
<td>61kB</td>
<td>1.0kB/s</td>
<td>Globus Notification <a href="mailto:no-reply@globus.org">no-reply@globus.org</a></td>
</tr>
<tr>
<td>2016-01-25 08:56:26Z</td>
<td>2016-01-25 09:02:00Z</td>
<td>ncsa#BlueWaters</td>
<td>grbot#H3ABioNetCentral</td>
<td>SUCCEEDED</td>
<td>33kB</td>
<td>1.0kB/s</td>
<td>Globus Notification <a href="mailto:no-reply@globus.org">no-reply@globus.org</a></td>
</tr>
<tr>
<td>2016-01-19 17:06:58Z</td>
<td>2016-01-29 08:46:05Z</td>
<td>ncsa#BlueWaters</td>
<td>grbot#H3ABioNetCentral</td>
<td>FAILED</td>
<td>7.2TB</td>
<td>73Mb/s</td>
<td>Globus Notification <a href="mailto:no-reply@globus.org">no-reply@globus.org</a></td>
</tr>
<tr>
<td>2016-01-15 14:59:49Z</td>
<td>2016-01-16 10:18:10Z</td>
<td>ncsa#BlueWaters</td>
<td>grbot#H3ABioNetCentral</td>
<td>SUCCEEDED</td>
<td>1.3TB</td>
<td>160Mb/s</td>
<td>Globus Notification <a href="mailto:no-reply@globus.org">no-reply@globus.org</a></td>
</tr>
<tr>
<td>2016-01-14 18:39:00Z</td>
<td>2016-01-14 23:09:32Z</td>
<td>ncsa#BlueWaters</td>
<td>ncsa#Nearline</td>
<td>SUCCEEDED</td>
<td>21TB</td>
<td>11Gb/s</td>
<td>Globus Notification <a href="mailto:no-reply@globus.org">no-reply@globus.org</a></td>
</tr>
<tr>
<td>2016-01-14 18:38:41Z</td>
<td>2016-01-14 20:43:42Z</td>
<td>ncsa#BlueWaters</td>
<td>ncsa#Nearline</td>
<td>SUCCEEDED</td>
<td>3.0TB</td>
<td>3.3Gb/s</td>
<td>Globus Notification <a href="mailto:no-reply@globus.org">no-reply@globus.org</a></td>
</tr>
<tr>
<td>2016-01-14 16:53:23Z</td>
<td>2016-01-14 19:50:18Z</td>
<td>ESnet test DTN at LBL</td>
<td>heinreedj#H3ABioNet</td>
<td>SUCCEEDED</td>
<td>500GB</td>
<td>400Mb/s</td>
<td>Globus Notification <a href="mailto:no-reply@globus.org">no-reply@globus.org</a></td>
</tr>
<tr>
<td>2016-01-14 10:36:57Z</td>
<td>2016-01-14 14:48:56Z</td>
<td>ESnet test DTN at CERN</td>
<td>heinreedj#H3ABioNet</td>
<td>SUCCEEDED</td>
<td>500GB</td>
<td>280Mb/s</td>
<td>Globus Notification <a href="mailto:no-reply@globus.org">no-reply@globus.org</a></td>
</tr>
<tr>
<td>2016-01-14 06:30:13Z</td>
<td>2016-01-14 06:44:11Z</td>
<td>ppanchal#hgs-c-xfer</td>
<td>ncsa#BlueWaters</td>
<td>SUCCEEDED</td>
<td>280GB</td>
<td>2.8Gb/s</td>
<td>Globus Notification <a href="mailto:no-reply@globus.org">no-reply@globus.org</a></td>
</tr>
<tr>
<td>2016-01-14 06:24:58Z</td>
<td>2016-01-14 06:25:02Z</td>
<td>ppanchal#hgs-c-xfer</td>
<td>ncsa#BlueWaters</td>
<td>SUCCEEDED</td>
<td>5.8kB</td>
<td>16kB/s</td>
<td>Globus Notification <a href="mailto:no-reply@globus.org">no-reply@globus.org</a></td>
</tr>
<tr>
<td>2016-01-14 17:31:34Z</td>
<td>2016-01-14 17:27:16Z</td>
<td>ncsa#BlueWaters</td>
<td>heinreedj#H3ABioNet</td>
<td>FAILED</td>
<td>1.4TB</td>
<td>70Mb/s</td>
<td>Globus Notification <a href="mailto:no-reply@globus.org">no-reply@globus.org</a></td>
</tr>
<tr>
<td>2016-01-11 12:28:34Z</td>
<td>2016-01-11 13:33:22Z</td>
<td>ppanchal#hgs-c-xfer</td>
<td>ncsa#BlueWaters</td>
<td>SUCCEEDED</td>
<td>2.7kB</td>
<td>0.0kB/s</td>
<td>Globus Notification <a href="mailto:no-reply@globus.org">no-reply@globus.org</a></td>
</tr>
<tr>
<td>2016-01-11 08:25:25Z</td>
<td>2016-01-11 20:42:08Z</td>
<td>ncsa#BlueWaters</td>
<td>grbot#H3ABioNetCentral</td>
<td>SUCCEEDED</td>
<td>560GB</td>
<td>110Mb/s</td>
<td>Globus Notification <a href="mailto:no-reply@globus.org">no-reply@globus.org</a></td>
</tr>
<tr>
<td>2016-01-11 08:25:04Z</td>
<td>2016-01-11 15:07:26Z</td>
<td>ncsa#BlueWaters</td>
<td>grbot#H3ABioNetCentral</td>
<td>SUCCEEDED</td>
<td>300GB</td>
<td>100Mb/s</td>
<td>Globus Notification <a href="mailto:no-reply@globus.org">no-reply@globus.org</a></td>
</tr>
<tr>
<td>2016-01-11 08:23:49Z</td>
<td>2016-01-11 16:03:44Z</td>
<td>ncsa#BlueWaters</td>
<td>grbot#H3ABioNetCentral</td>
<td>SUCCEEDED</td>
<td>520GB</td>
<td>160Mb/s</td>
<td>Globus Notification <a href="mailto:no-reply@globus.org">no-reply@globus.org</a></td>
</tr>
<tr>
<td>2016-01-11 08:23:26Z</td>
<td>2016-01-11 18:12:39Z</td>
<td>ncsa#BlueWaters</td>
<td>grbot#H3ABioNetCentral</td>
<td>SUCCEEDED</td>
<td>610GB</td>
<td>140Mb/s</td>
<td>Globus Notification <a href="mailto:no-reply@globus.org">no-reply@globus.org</a></td>
</tr>
</tbody>
</table>
## Data transfer

### UDP v/s TCP

<table>
<thead>
<tr>
<th>Characteristics/Description</th>
<th>UDP</th>
<th>TCP</th>
</tr>
</thead>
<tbody>
<tr>
<td>General Description</td>
<td>Simple High speed low functionality “wrapper” that interface applications to the network layer and does little else</td>
<td>Full-featured protocol that allows applications to send data reliably without worrying about network layer issues.</td>
</tr>
<tr>
<td>Protocol connection Setup</td>
<td>Connection less data is sent without setup</td>
<td>Connection-oriented; Connection must be Established prior to transmission.</td>
</tr>
<tr>
<td>Data interface to application</td>
<td>Message base-based is sent in discrete packages by the application.</td>
<td>Stream-based; data is sent by the application with no particular structure</td>
</tr>
<tr>
<td>Reliability and Acknowledgements</td>
<td>Unreliable best-effort delivery without acknowledgements</td>
<td>Reliable delivery of message all data is acknowledged.</td>
</tr>
<tr>
<td>Retransmissions</td>
<td>Not performed. Application must detect lost data and retransmit if needed.</td>
<td>Delivery of all data is managed, and lost data is retransmitted automatically.</td>
</tr>
<tr>
<td>Features Provided to Manage flow of Data</td>
<td>None</td>
<td>Flow control using sliding windows; window size adjustment heuristics; congestion avoidance algorithms</td>
</tr>
<tr>
<td>Overhead</td>
<td>Very Low</td>
<td>Low, but higher than UDP</td>
</tr>
<tr>
<td>Transmission speed</td>
<td>Very High</td>
<td>High but not as high as UDP</td>
</tr>
<tr>
<td>Data Quantity Suitability</td>
<td>Small to moderate amounts of data.</td>
<td>Small to very large amounts of data.</td>
</tr>
</tbody>
</table>
Data transfer – Globus online

1. User initiates transfer request
2. Globus moves and syncs files
3. Globus notifies user

Data Source

Data Destination
Data transfer – Globus online

Activity

heinedej#H3ABioNet to grbot#H3ABioNetCentral
warning: connection failed; see event log
Data transfer – Globus online
Data transfer – Globus online
Data transfer – Globus online

![Globus interface showing data transfer activity with a warning for a connection failure]

Activity

Overview

Event Log

- **Warning:** Connection failed; see event log

<table>
<thead>
<tr>
<th>Date/Time</th>
<th>Event Description</th>
<th>Link to View Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>2016-05-12 08:16 am</td>
<td>The operation was started or restarted</td>
<td></td>
</tr>
<tr>
<td>2016-05-12 08:15 am</td>
<td>Could not connect</td>
<td></td>
</tr>
<tr>
<td>2016-05-12 08:14 am</td>
<td>The operation was started or restarted</td>
<td></td>
</tr>
<tr>
<td>2016-05-12 08:14 am</td>
<td>The operation succeeded</td>
<td></td>
</tr>
<tr>
<td>2016-05-12 08:13 am</td>
<td>The operation was started or restarted</td>
<td></td>
</tr>
</tbody>
</table>
Data transfer – Globus online

Advantages:
- robust
- efficient
- faster

Disadvantages:
- complex setup
- non-standard firewall configuration
- can hog bandwidth
Storage

• Types of storage
  – Volatile
    • RAM
    • variables
    • cache
  – Persistent
    • disk
    • RAID
    • tape
Storage

RAID 0

RAID 1

RAID 5

Source: https://en.wikipedia.org/wiki/Standard_RAID_levels
Storage - iRODS

• Advantages
  – Associated metadata
  – Rules oriented
  – Built in provenance

• Disadvantages
  – Complex setup
  – Rules infrastructure
  – Integration with existing tools
Security

• The most secure machine is a disconnected, physically secured machine that is powered off.

• Goals
  – Confidentiality
  – Authentication
  – Integrity
  – Non-repudiation

• Security is a trade-off
Security – threat modelling

– What do you have that is valuable? (assets)
– Why would attackers want to disrupt (motivation)
– Where can they attack (entry points)
– How would they attack (threats)
– Cost to protect? (threat ranking)
– Which threats will you fight and how (mitigation)
Security – motives

– Obtain asset (personal info, genetic data)
– Use resources (e.g. CPU, network)
– Recognition
– Thrill
– Mistakes
Security – types of attack

• Denial of service
• Exploit configuration errors
• Exploit application bugs
• Social engineering
Security – calculating risk

• Exploitability
  – Likelihood of vulnerability being exploited

• Damage potential
  – Consequences of breach

• Asset value
  – Cost of protecting, recovery
Security - cryptography

• Terms
  – Cleartext
  – Cipher
  – Ciphertext

• Goals
  – Confidentiality
  – Authentication
  – Integrity
  – Non-repudiation
Security - principles

– Compartamentalize
– Principle of least privilege
– Defense in depth
– Don’t volunteer information
– Fail safely
– Weakest link
– Simplicity
Security - cryptography

• Symmetric encryption
  – requires a shared “secret key”
Security - cryptography

• Asymmetric encryption (public key encryption)
  – Key pairs
  – No shared secret required
Security - cryptography

• Tools
  – GPG

• Related
  – Public Key infrastructure
  – Digital certificates
  – Certificate authorities
Conclusion

• Security is hard
• Resources:
  – Bruce Schneier
  – http://security.stackexchange.com
  – https://twitter.com/thegrugq