Never Underestimate the Bandwidth of a Truck Full of Hard Drives Hurtling Down the Highway

(apologies to Andrew S. Tanenbaum)

Tim Haverland, NOAA Fisheries Office of Science and Technology
What we’ll cover

The history of “sneakernets”

Bringing data and computing together

Physically shipping data to the cloud
Some history ...
Wi-Fly?

Average speed: 50 mph

Max distance: 400 miles

Distance to nearest AWS data center

By car: 33 miles (49 mins)

Pigeon: 23 miles (28 mins)

Two hours reading/writing

Effective transfer rate: 903 Mb/s!
Never underestimate ...

... the bandwidth of a station wagon full of tapes hurtling down the highway

- *Computer Networks*, 3rd ed., p. 83. (paraphrasing Dr. Warren Jackson, Director, University of Toronto Computing Services (UTCS) circa 1985)
Cuba’s *Paquete Seminal* (Weekly Packet)

Cuba has one wifi hotspot for every 35,000 Cubans

El Paquete is a terabyte of movies, songs, text and TV shows

Distributed on hard drives by bus, foot, horseback to over 3 million Cubans

Leaves Havana at 3am .... The rest of Cuba by 8am

Nestor Siré and Julia Weist
Storage density vs network bandwidth

Storage density increases by 15% per year ... or more?

NOAA network bandwidth ... ?

<table>
<thead>
<tr>
<th></th>
<th>Hard disk</th>
<th>Flash memory</th>
<th>Bacterial DNA</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Read-write speed</strong></td>
<td>~3,000–5,000</td>
<td>~100</td>
<td>&lt;100</td>
</tr>
<tr>
<td>(µs per bit)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Data retention</strong></td>
<td>&gt;10</td>
<td>&gt;10</td>
<td>&gt;100</td>
</tr>
<tr>
<td>(years)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Power usage</strong></td>
<td>~0.04</td>
<td>~0.01–0.04</td>
<td>&lt;10⁻¹⁶</td>
</tr>
<tr>
<td>(watts per gigabyte)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Data density</strong></td>
<td>~10¹³</td>
<td>~10¹⁶</td>
<td>~10¹⁹</td>
</tr>
<tr>
<td>(bits per cm³)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**STORAGE LIMITS**

Estimates based on bacterial genetics suggest that digital DNA could one day rival or exceed today’s storage technology.
<table>
<thead>
<tr>
<th></th>
<th>Physical Transfer</th>
<th>Physical / Online Transfer</th>
<th>Online Transfer</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1 Mbps</td>
<td>10 Mbps</td>
<td>100 Mbps</td>
</tr>
<tr>
<td>1 GB</td>
<td>3 hours</td>
<td>18 minutes</td>
<td>2 minutes</td>
</tr>
<tr>
<td>10 GB</td>
<td>30 hours</td>
<td>3 hours</td>
<td>18 minutes</td>
</tr>
<tr>
<td>100 GB</td>
<td>12 days</td>
<td>30 hours</td>
<td>3 hours</td>
</tr>
<tr>
<td>1 TB</td>
<td>124 days</td>
<td>12 days</td>
<td>30 hours</td>
</tr>
<tr>
<td>10 TB</td>
<td>3 years</td>
<td>124 days</td>
<td>12 days</td>
</tr>
<tr>
<td>100 TB</td>
<td>34 years</td>
<td>3 years</td>
<td>124 days</td>
</tr>
<tr>
<td>1 PB</td>
<td>340 years</td>
<td>34 years</td>
<td>3 years</td>
</tr>
<tr>
<td>10 PB</td>
<td>3,404 years</td>
<td>340 years</td>
<td>34 years</td>
</tr>
<tr>
<td>100 PB</td>
<td>34,048 years</td>
<td>3,404 years</td>
<td>34 years</td>
</tr>
</tbody>
</table>
Bringing Muhammad to the Mountain

Bring data close to computing

Unlocks new analytical capabilities (example of NEXRAD)

Reduces expensive and time consuming data transfers

Major cloud providers all have physical data shipping mechanisms
Amazon Snowball

80TB in about 10 days to AWS

$250 not including shipping
Amazon Snowmobile!

Up to 100 Petabytes (PB)
Transfer to cloud in “a few weeks”
330Gb/s transfer speed 100 PB @ 4 weeks

$0.005/GB/month
$500,000
Google Transfer Appliance

100-200 TB

$800

480 TB - 1PB

$2,700

1 PB transfer

Typical Network
100Mbps

1095 Days
(3 Years)

Transfer Appliance

43 Days
(Way Faster)
IBM Cloud Mass Data Migration

120TB

“As little as 7-days to the cloud”

$395
Azure

Accepts shipped SATA HDDs and SDDs

10 drives max in an import job

$80/drive not including shipping

Copy files as normal

Drives are shipped back to you (!)

Can also do an export to get data *out* of Azure and onto drives
In conclusion ...

Our future networks will have trouble keeping up with data volumes

... and with storage density improvements

Physical data shipping will become even more important
Questions?

tim.haverland@noaa.gov

301-427-8137