HPSS at Indiana University
Site Report
History of HPSS at IU

- We have been using HPSS for 20 years
- Replicated between campuses 60 miles apart
- Initially used HPSS-DFS v4
- Fixed DFS problems by getting rid of it.
• Gatewayed Services
  – Globus
  – SFTP
  – R/O Samba
• Hsigwd on disk movers
Recent Updates

- Upgraded movers and gateways
- Tape libraries type
- Drives: 3592-E08
  - 28 in Indy
  - 32 in Bloomington
- Disk cache DDN 12KE (Bloomington) and 7990 (Indy)
New Features

- DSI Globus
- QED (Quota Editor)
- Usage analysis with Splunk
- HPSS 7.4.3 -> 7.5.3 upgrade
DSI Globus spanning 4 servers

- **Pros**
  - Easy interface to other Globus endpoints, like our own DC2 and Geode (planned)
  - Third party transfer, set-and-forget
  - No longer depends on VFS/FUSE
- **Cons**
  - Users who dump thousands of files at once
  - Speed isn't optimal (roughly 60-80 MB/s)
QED (Quota EDitor) by Jeff Russ

- Uses GK to enforce
- Limits both by data stored and number of files

```
[root@newcsb1 ~]# qed
usage: qed -a -d data_quota -f file_quita -l -o -v one_or_more_users
[root@newcsb1 ~]# qed ecgarris

<table>
<thead>
<tr>
<th>User</th>
<th>File Quota</th>
<th>Data Quota in GB</th>
<th>Files</th>
<th>Data in GB</th>
</tr>
</thead>
<tbody>
<tr>
<td>ecgarris</td>
<td>25,000</td>
<td>50,000</td>
<td>490</td>
<td>392</td>
</tr>
</tbody>
</table>

[root@newcsb1 ~]# qed -a | head

<table>
<thead>
<tr>
<th>User</th>
<th>File Quota</th>
<th>Data Quota in GB</th>
<th>Files</th>
<th>Data in GB</th>
</tr>
</thead>
<tbody>
<tr>
<td>root</td>
<td>200,000</td>
<td>1,000</td>
<td>177,256</td>
<td>657</td>
</tr>
<tr>
<td>luehring</td>
<td>25,000</td>
<td>50,000</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>russ</td>
<td>89,458</td>
<td>50,000</td>
<td>64,473</td>
<td>18,731</td>
</tr>
<tr>
<td>detran</td>
<td>25,000</td>
<td>50,000</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>shburks</td>
<td>25,000</td>
<td>50,000</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>lwm</td>
<td>25,000</td>
<td>50,000</td>
<td>318</td>
<td>645</td>
</tr>
<tr>
<td>sweeny</td>
<td>25,000</td>
<td>50,000</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>basset</td>
<td>25,000</td>
<td>50,000</td>
<td>2</td>
<td>0</td>
</tr>
</tbody>
</table>
```
Usage analysis (Splunk stats)

- HSI/HTAR – two years of data
  - 19 million connections
  - 97% from within IU networks
- Sftp – one year’s data
  - 50,000 connections
  - 91% of connections from within IU networks
- Globus – Not a lot of usage yet, but expect it to grow
• Typical weekly usage
## Using Existing Tape Inventory

### Burn-rate/day
- **8.93**

### Days remaining
- **623.84**

### Projected date
- **6/26/21**
**HIPAA Usage**
- Known users
- All transactions logged

### HIPAA_all

**All users flagged with PHI**

<table>
<thead>
<tr>
<th>Time</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>2018-10-24T21:34:03-0400</td>
<td>Opened at 10242018 21:34:03, Closed at 10242018 21:34:03, User: 74, Reqld:</td>
</tr>
</tbody>
</table>
HPSS 7.5.3 upgrade

- Problems converting metadata
- 20 year old system, we have cruft and corrupt metadata from HPSS v4 that has needed sorting out.
- Upgrade while live has caused numerous challenges
MDPI

• Significant user of our HPSS
• Workflow of checksum, upload, purge, stage, checksum, external copy
• HPSS 7.5 has end to end data integrity so it should lessen this workflow
Backup Service Testing

- Bareos (Bacula fork) on top of FUSE hpssfs
- Looks like a local directory to Bareos
- One department uses Veeam over gatewayed sshfs
Looking forward

- Refreshing core servers
- Moving to new tape libraries, our libraries are 10 years old
Conclusion

• Looking to hear experiences with other tape libraries, flash drives for metadata and backup solutions.