

# Indiana University's HPSS – A Site Report

Anurag Shankar <ashankar@indiana.edu>

Manager, Distributed Storage Services  
University Information Technology Service  
Indiana University

HPSS User Forum, June 18-20, 2002

[http://storage.iu.edu/presentations/iu\\_rept\\_huf02.ppt](http://storage.iu.edu/presentations/iu_rept_huf02.ppt)

## Contributors : Past and Present

- Dave Vernon (now at Cornell U.)
- Gustav Meglicki
- Leigh Grundhoefer
- Jeff Russ
- Chris Hemmerich
- Haichuan Yang
- Chris Garrison
- Andy Arenson

## Outline

- History
- Statistics
- HPSS Use at IU
- Configuration
- Unique/Noteworthy Features
- Issues/Concerns
- Future
- Conclusion

## History

- IU started looking at storage trends in 1998. We went through a RFP process in 1999 to build a massive data storage system.
  - Main hierarchical storage management (HSM) software choices included HPSS (High Performance Storage System), LSC's SamFS, and UniTree.
  - HPSS chosen since it seemed the most likely technology to survive long term (it's a collaboratively developed, not a niche vendor supplied solution).

## History...

- IU joined the HPSS collaboration in early 1999. IBM chosen as the hardware vendor. Base system in production at Bloomington in June 1999.
- The DFS front-end to HPSS in production in August 1999.
- Remote movers across a WAN in production at Indianapolis in October 2000.
- Gatewayed HPSS for the masses available in June 2002.

## Statistics

- Users : 1,000
  - Data stored : 65 TB
  - Files : 4,279,875
  - Data transfer : 50 GB – 1TB per day
  - Data storage rate : +20 TB per year
  - Data storage pattern : > 98% writes
- (Detailed stats at <http://www.indiana.edu/~dssg/stats.html>)

## Statistics...

- Uptime : 98.4%
- Downtime details :
  - 100 hours (regularly scheduled)
  - 40 hours (unscheduled)
- # of problems : 46

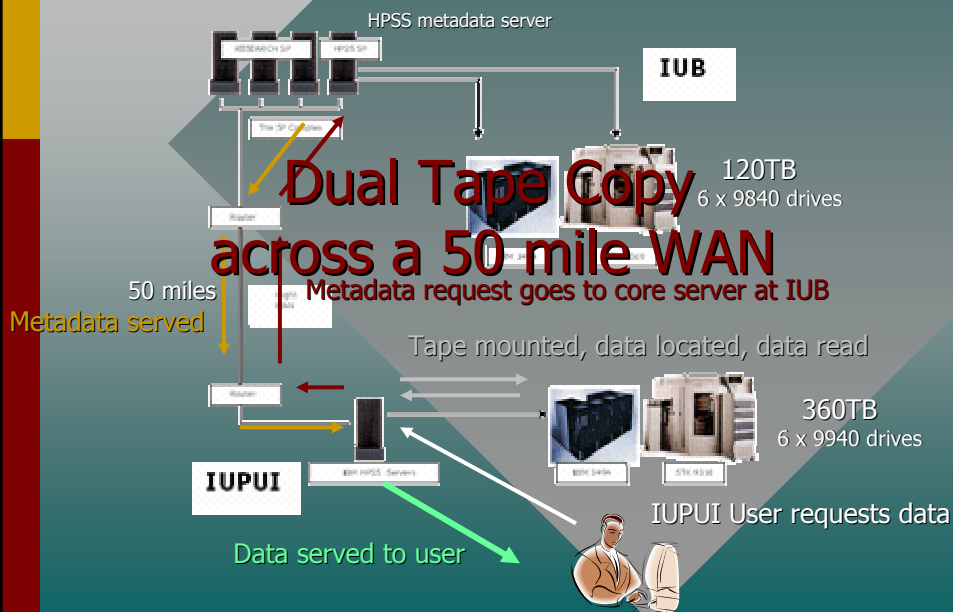
## HPSS Use at IU

- Research – traditional fields (astronomy, physics, chemistry, geology, etc.), but popular now with hitherto unrepresented fields at IU (genomics, biology, theater, fine arts, business, library IS, history, etc.).
- Medical – at the IU School of Medicine.
- Administrative – student records, scanned documents, HR data, etc.
- Digital libraries –images, audio, video, scanned books, etc.
- General –workstation backups, etc.

## Configuration (Hardware)

- Core servers run on a wide "Silver" IBM SP node at Bloomington.
- Tape/disk movers and migrating DFS servers configured on 10xIBM SP nodes at Bloomington, and on a IBM H70 and 3xB80 nodes in Indianapolis, across a 50 miles, across a 50-mile WAN.

### Distributed HPSS between Bloomington and Indianapolis



## Configuration (Software)

- At HPSS 4.1.1.4.
- At AIX 4.3.2, DCE 2.2 clients, DFS 2.2 servers, DCE 2.2 core servers.
- Using Samba, Netatalk, Apache, s-FTP as gateways, to provide access to the masses.

## Configuration (Software)...

HPSS accessed by users via:

- FTP
  - Parallel FTP
  - DFS
  - HSI
  - Samba/Netatalk/s-FTP
  - Apache (WWW)
- Slow but quite popular
-

## Configuration (Software)...

- HPSS accessible without a client as a mapped drive letter under Windows (9x, NT4, 2000, XP). → \\mdss.iu.edu\- Accessible without a client as an Appleshare IP volume under MacOS (8, 9, X). → mdss.iu.edu
- Accessible via smbclient/smbfs from Unix/Linux.
- Secure web access from any browser.  
- <http://mdss.iu.edu/<username>>

## Unique/Noteworthy Features

- IU making a high-end, HSM system accessible to the masses.
- Establishing remote/satellite HPSS data movers.
- Establishing remote dual tape copy over a WAN, making HPSS data disaster proof.
- Handling novices to high-end users with a seamlessly woven storage, DFS-enabled fabric.
- Provided service with a mere 1 FTE (2 now).

## Issues/Concerns

- Small files – looking at the SRB. Packed fileset?
- DFS disk cache issues – may go away along with DFS. How to replace DFS?
- Interfacing HPSS with Windows – Microsoft W2K server possibilities?
- Weekend, after-hours emergency support?

## Future

- Upgrade to HPSS 4.3 this summer.
- Plan a future without DCE/DFS.
- Implement remote dual tape copy across a WAN.
- Make Samba/Netatalk/Apache HPSS-aware.

## Conclusions

- HPSS is working well at IU.
- HPSS is making a fundamental difference in the IU researcher's ability to get funding, to get work done more efficiently.
- We need a distributed, file system bridge for HPSS to be appealing to the masses.

For more info, visit <http://storage.iu.edu/>.