Storage@IU

Supercomputing for Everyone Series: Faster Work, Safer Storage
Feb 17-18, 2016

Research Technologies - Systems
Indiana University
Overview of presentation

1. Data Workflow
   - Thinking about the lifecycle and types of data you use and create

2. Storage Resources
   - How to most efficiently store and use data
   - Policies, best practices, and methods for optimizing performance

3. Getting your data in and out of storage systems

Questions welcome any time!

There will also be time at the end for discussion.
Consider the types of data you may have

Program code
Documentation

Scientific data
Computational data
Intermediate steps

Output
Results of computation
Simple Workflow

Input instructions

Read, compute, write

Archive results
## Data Storage Requirements

<table>
<thead>
<tr>
<th>Source Code &amp; Documents</th>
<th>Computational Data</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>– Easily accessible</td>
<td>– Fast</td>
<td>– Store safely for long time</td>
</tr>
<tr>
<td>– Backed-up</td>
<td>– Large capacity</td>
<td>– Potentially very large</td>
</tr>
<tr>
<td></td>
<td>– Collaboration</td>
<td>– Sharing</td>
</tr>
</tbody>
</table>

- Easily accessible
- Backed-up
- Fast
- Large capacity
- Collaboration
- Store safely for long time
- Potentially very large
- Sharing
Home Directory
- Default location
- For small files
- Data is backed up

/N/u/username/BigRed2

Data Capacitor II (DC2)
- Large capacity
- High throughput
- For compute data
- Data is not backed up

/N/dc2/

Scholarly Data Archive (SDA)
- Disk to tape archiving
- Distributed copies
- For long term storage
Storage Resource Analogies

• Home Directories –
  • Lab Notebook
    • Relatively small
    • Instructions for experiment

• Data Capacitor II –
  • Lab Whiteboard
    • Large and quick workspace
    • Erased over time

• Scholarly Data Archive –
  • Lab Library
    • Very large space
    • Safe place for critical data
Home Directory – The Lab Notebook

• Default storage location for your account
  – Available as soon as you log in

• The place to store source code, shell scripts, and other small files

• Not meant for computational data!
  • Do not compute against data in home directories!
    – Use Data Capacitor II for computational data

• More information:
  https://kb.iu.edu/d/avkm
Home Directory – Shared Across Systems

- Home directories space (100GB quota) shared across RT systems

<table>
<thead>
<tr>
<th>System</th>
<th>Path</th>
</tr>
</thead>
<tbody>
<tr>
<td>Big Red 2</td>
<td>/N/u/username/BigRed2</td>
</tr>
<tr>
<td>Mason</td>
<td>/N/u/username/Mason</td>
</tr>
<tr>
<td>Karst</td>
<td>/N/u/username/Karst</td>
</tr>
</tbody>
</table>

username@login1:~> pwd
/N/u/username/BigRed2
username@login1:~> cd ..
username@login1:/N/u/username>
ls
BigRed2  Mason  Karst
Data Capacitor II

Compute against data
Data Capacitor II – The Lab Whiteboard

• Parallel high-speed storage based on Lustre file system
  • 3.5 PB total size, ≥40 GB/s throughput to BR2

• Store application’s input and output data
  – Intended as a temporary workspace for computation
    » not for indefinite storage of data
  – DC2 is not backed up

• Available on IU’s HPC resources
  – Big Red II, Mason, Karst

• More information available on the Knowledge Base
  • https://kb.iu.edu/d/avvh
Data Capacitor II – Scratch Directories

• A scratch directory is available to every HPC user
  • Path to your scratch space is /N/dc2/scratch/username/
  • Intended as temporary workspace for your data, not for sharing
  • Files not accessed in 60 days may be purged
  • The name “scratch” is from the phrase “scratch paper”
    – A piece of paper used while performing calculations
    – Implies an impermanence to the data
Data Capacitor II – Project Directories

- Project directories for users, groups, or labs with special needs
  - Makes it possible to share data amongst users (Unix groups)
  - Files not accessed in **180 days** may be purged
  - Project space can be applied for by submitting application at this URL
  - Longer storage time than scratch space, but still not forever
Data Capacitor II – Purging Old Data

• Administrators of Data Capacitor II routinely purge old data
  • Data not accessed in a certain amount of time will be deleted
    – scratch = 60 days, project = 180 days
• You will be notified before and after any action is taken against your data
  • An email will be sent to you listing the eligible files
  • A file will be place in the root of your scratch or project directory
    – Will-Purge-These-uid648424-username-Files-On-2014-06-30.txt
• You will have seven (7) days to take action
• Afterwards, a file listing the actions taken will be created

See https://kb.iu.edu/d/avvh#usage
Data Capacitor II – Find Old Files

- You can be proactive about managing your data to prevent purging
  - Use this command to list the files in your directory sorted by age
    - May take a while, it depends on the number of files you have

```
find /N/dc2/scratch/ <username> -type f -exec stat --format="%n %x" '{}' \; | sort -k2,3
```
Data Capacitor II – Space Quota

• There is no strict limit on the amount of space you can use
  • Space available for all users varies depending on system use
    – ‘df –h /N/dc2/scratch’ gives you current space available
    – du –hc /N/dc2/scratch/ <username> gives you the amount of space that you are using
  • Please don’t use any more space than strictly necessary
    – Data Capacitor II is a shared resource
      – intended for computation
Data Capacitor II – HIPAA and ePHI

- DC2 is HIPAA aligned, but you are responsible for ensuring the privacy and security of ePHI data (see https://kb.iu.edu/d/ayzm)

- Technical safeguards
  - Set directory permissions to restrict read and write access
    - The most secure method is to allow access only to you

    ```
    username@login1:/N/dc2/scratch/username> chmod 700 ephi_file
    username@login1:/N/dc2/scratch/username> ls -l ephi_file
    -rwx------ 1 username uits 0 Jan 23 14:25 ephi_file
    ```

  - Use `umask` to ensure all new files are created with safe permissions
    - Add `umask 077` to shell profile
    - See https://kb.iu.edu/d/acge
Data Capacitor II – Job Scheduling

- Specify DC2 as a requirement for your batch job
  - add the “dc2” file system property to the nodes directive in your in
    your TORQUE job script
- For example, if your job requires two nodes, thirty two processors per node,
  and the Data Capacitor II file system (/N/dc2), the resource specification line
  in your TORQUE job script would look like:

```bash
#PBS -l nodes=2:ppn=32:dc2
```

- Specifying the dc2 property in your script directs TORQUE to dispatch your
  job to only those compute nodes with the Data Capacitor II file system
  mounted. If DC2 is down, your job won’t run.
- More information at: https://kb.iu.edu/d/baht
Data Capacitor II – Hardware Details

- 1,680 total 3 TB SATA drives – 3.5 PB formatted
- High Throughput
  - ~20 GB/s via Ethernet
  - >40 GB/s via InfiniBand
  - Ethernet systems: Karst, Mason
  - InfiniBand systems: Big Red II
Data Capacitor II – Usage

Lustre is designed for high-speed data access, not for metadata speed

- Tips to improve performance:
  - Avoid more than 10K files in one directory
    - Separate input, output, final results, and delete unneeded data
  - Limit the amount of metadata actions you perform
    - Reduce file and directory operations, `stat-ing` files
  - Data Striping
    - allows you to control how data is written
    - [https://www.nics.tennessee.edu/computing-resources/file-systems/io-lustre-tips](https://www.nics.tennessee.edu/computing-resources/file-systems/io-lustre-tips)
Striping Example

File 1 - 1MB
File 2 - 4MB
File 3 - 7MB

1 OSS
4 Stripes
1MB Stripe Size
Data Capacitor II – Reporting Issues

If you encounter any problems using Data Capacitor II, please include these details when reporting the issue:

• Data and time event occurred
• Which system your job was running on
• The directory being used
• A brief description of what was happening when the issue occurred

All Data Capacitor II issues should be reported to hpfs-admin@iu.edu
Scholarly Data Archive - The Lab Library
Scholarly Data Archive

- Massive near-line and archival data storage
  - Disk cache front end – ~1800TB
  - Magnetic tape storage – ~15 PB
- Data migrates from disk to tape over time
- Data integrity
  - Replicated at IUB and IUPUI
  - Checksums and error detection
Scholarly Data Archive – Details

• Account can be applied for easily
  • [https://itaccounts.iu.edu/](https://itaccounts.iu.edu/)
• Default quota is 50 TB
  • Replicated copy of data is not counted
  • Additional storage is available
• HIPAA aligned but you must secure the data
  • [https://kb.iu.edu/d/ayzm](https://kb.iu.edu/d/ayzm)
• Replicated copies of data, but no backups
• Group or department accounts are available
• Data can be shared with Access Control Lists
• More information available on the Knowledge Base
  • [https://kb.iu.edu/d/aiyi](https://kb.iu.edu/d/aiyi)
Scholarly Data Archive – Usage

• Best Uses
  • Files larger than 1MB & smaller than 10TB
  • Archive files that rarely change

• Poor Uses
  • Small files (archive first with tar, zip, etc.)
  • Files that will frequently change
  • Do not edit files in place
    – If you need to edit: Copy -> Edit -> Re-upload
Scholarly Data Archive – Classes of Service

- Classes of service determine how data is cached and stored.
- There are 5 classes of service based on file size: 1, 2, 3, 4, and 13. All except for 13 are dual-copy at IU Bloomington & IUPUI.
- If your data is large and stored elsewhere or easily regenerated, use class 13 to save system space.
- System usually picks automatically, but important to understand basics.
- More information at https://kb.iu.edu/d/auli
Scholarly Data Archive – Helpful Tips

• Data stored on the SDA can be kept for a long time
  • So long that you might even forget
    – Or the people who did know have left
• Do your future self a favor and document the data
  • Create a manifest or annotation of the data
  • Keep it at the top of your storage directory, and keep it up to date
• Contact the Research Storage team if you have any SDA questions
  • https://mailform.kb.iu.edu/email.php?cid=59
Transferring Data In and Out of RT Storage
Preparing to Transfer Data

• It is recommended to bundle your data before transferring
  - Easier to manage a single file
  - Preserves layout, permissions
  - Transferring large files is often faster than many small files

```
username@login1:/N/dc2/scratch/username> ls
input  output  results
username@login1:/N/dc2/scratch/username> tar -cvf archive.tar input/ output/ results/
input/
output/
results/
username@login1:/N/dc2/scratch/username> ls
archive.tar input output results
```
Getting data in and out of Home Directories

- scp is the easiest way to get data in and out of your home directory
  - Secure
  - sftp and rsync over ssh are also good options

$ scp archive.tar bigred2.uits.iu.edu:~
Getting data in and out of Data Capacitor II

- The IU Cyberinfrastructure Gateway allows you to transfer data between your machine and Data Capacitor II
  - IU CI Gateway information:
    - [https://kb.iu.edu/d/bdfo](https://kb.iu.edu/d/bdfo)
  - Transferring data with CI Gateway to DC2:
    - [https://kb.iu.edu/d/bdq](https://kb.iu.edu/d/bdq)
- The IU CI Gateway uses Globus Online
  - A parallel transfer tool which requires software to be installed
    - The endpoint is iu#dc2
  - Follow the instructions in the KB article to request a DC2 account
  - `scp/rsync/sftp` available as well
Getting data in and out of Scholarly Data Archive

- Fast access
  - hsi and htar command line tools
    - To use HSI on HPC systems, you must load the HPSS module
      - module load hpss
  - GridFTP clients
  - Kerberized FTP
  - GlobusOnline – also available through the IU CI Gateway

- Convenience protocols
  - Web access via browser
  - sftp
  - Mount to desktop via CIFS/Samba (mapped drive)

- Knowledge Base article on SDA access
  - https://kb.iu.edu/d/aiyr
Data Transfer Considerations

- Big Red II, Mason, Karst login nodes do not enforce a time limit on data transfer tools
  - scp, sftp, hsi, htar, wget, curl, etc.

- Recommend putting your data into the Scholarly Data Archive first
  - Then use command line tools to pull from SDA into DC2, Home Directories
  - Many ways to access the SDA, robust permissions
Lab: Uploading Data To SDA with SDA Web Interface

- Find or create a sample file to work with (text editor, etc.)
- Navigate with web browser to https://www.sdarchive.iu.edu
- Click “Upload File” at the top left
- In the Upload File dialog box, click Choose File or Browse to access your local computer's file system and find the file you want to upload. Click the file to select it.
- To start the upload, click Upload File. To cancel the upload, click Close.
- More information on SDA web interface is available at: https://kb.iu.edu/d/auxl
Lab: Transferring data from SDA to DC2 with HSI

- Login to bigred2.uits.iu.edu
- “cd /N/dc2/scratch/<username>”
- “module load hpss”
- “hsi” (then enter network-id and password)
- “ls” (file you uploaded should be shown)
- “get <filename>” (can also use “put” to upload files)
- “quit”

More information on SDA HSI access at: https://kb.iu.edu/d/avdb

HSI documentation at: http://www.mgleicher.us/index.html/hsi/
Lab: Transferring data from DC2 to SDA with SFTP

- Login to bigred2.uits.iu.edu
- “cd /N/dc2/scratch/<username>”
- “cp <filename> test_file”
- “sftp sftp.sdarchive.iu.edu” (will be prompted for password)
- “put test_file”
- “ls” (should see test_file as well as original file uploaded)
- “rm test_file” & “rm <filename>”
- “quit”
- More information on SDA SFTP access at: https://kb.iu.edu/d/avax
Other RT Storage Resources

• Focus of this presentation was basics of available research storage.
  • There are more storage options available
    – Data Capacitor WAN (DC-WAN)
      – Lustre over the wide area network, share at high speed
        – https://kb.iu.edu/d/avvh
    – Research File System (RFS)
      – 100GB centralized disk storage
      – Many access methods
        – https://kb.iu.edu/d/aroz
System Outages & Maintenance

• Problems with the storage systems posted via IT Notices
  • [http://itnotices.iu.edu/](http://itnotices.iu.edu/)

• Data Capacitor II regularly scheduled system maintenance
  • First Tuesday of every month (7am-7pm)
  • Join the maintenance mailing list to be notified
    – [https://kb.iu.edu/d/avvh#support](https://kb.iu.edu/d/avvh#support)

• Scholarly Data Archive regularly scheduled system maintenance
  • Weekly on Sunday (7am-10am)
Questions?
Attributions

• "Blank whiteboard" by BrokenSphere - Own work. Licensed under CC BY-SA 3.0 via Commons - https://commons.wikimedia.org/wiki/File:Blank_whiteboard.JPG#/media/File:Blank_whiteboard.JPG

• "Otto Hahn's notebook 1938 - Deutsches Museum - Munich" by J Brew - originally posted to Flickr as Nuclear fission Deutsches Museum. Licensed under CC BY-SA 2.0 via Commons - https://commons.wikimedia.org/wiki/File:Otto_Hahn%27s_notebook_1938_-_Deutsches_Museum_-_Munich.jpg#/media/File:Otto_Hahn%27s_notebook_1938_-_Deutsches_Museum_-_Munich.jpg