Overview

- IQSS data
  - What?
  - Where?
  - How?
  - Encryption

- Backup policies
  - When?

- Cloud storage / backup

- Questions
What we store

• Data volumes are allocated for specific types of data:
  • Murray Research Archive
  • Web server files
    – DVN
    – CGA WorldMap / AfricaMap
  • RCE storage
    – Home directories
    – Project directories (backed up and not backed up)
  • CGIS accounts
    – Home and group directories
    – Department shares
  • FTP mirrors of open-source files
  • Tools data
    – e.g., source (code) control data (CVS)
  • Mail
Where we store data

User (applications)

Server

Internal drives

...

Server

Internal drives

NetApp filer

Head 1

Head 2

Head 3

NetApp filer

Tape drive

DR site (60 Ox)
Protocol cheat sheet

- HTTP = HyperText Transfer Protocol (web access)
- SFTP = Secure File Transfer Protocol
- NFS = Network File System (network file access)
- CIFS = Common Internet File System (MS version of NFS)
- NDMP = Network Data Management Protocol (backups)
- NetVault = proprietary backup software
- SnapMirror = NetApp file replication technology
How we store data

User (applications)

Server
Internal drives

...(three servers shown)

Server
Internal drives

Server
Internal drives

NetVault

NetApp filer

NetVault

NetVault

NetVault

Tape drive

HTTP
SFTP

HTTP
SFTP

CIFS

NFS

NFS

NDMP

SnapMirror

DR site (60 Ox)
Backup encryption

• Backups are encrypted
  • Using LTO4 format, LTO3 does not support encryption

• Encryption keys are stored on local storage on servers
  • Redundant local disks store redundant copies
  • If local server has catastrophic failure, encryption keys are stored on IronKey (secure USB drive)

• Encrypted data is transferred from servers to Tape Backup Device
Backup encryption

User (applications)

HTTP SFTP

IronKey (security Keys)

Internal drives

Server

HTTP SFTP

CIFS

NetVault

NFS

NetVault

NDMP

Tape drive

NetApp filer

Head 1

Head 2

Head 3

DR site (60 Ox)

NetVault

SnapMirror

NetApp filer

Internal drives

Server

Internal drives
Backup frequency

- Murray Research Archive
  - Incremental backup weekly
  - Full backup every six months
    - Contractually required

- Other storage
  - DR (Disaster Recovery) sync every 15 minutes
  - Incremental backup daily
  - Full backup monthly
How often we back up

User (applications)

HTTP SFTP
IronKey (security Keys)
HTTP SFTP
CIFS

NetApp filer

Server

Internal drives

NetVault @ 1 day

Tape drive

NetVault @ 1 day

NDMP @ 1 day (incremental) @ 1 month (full)

SnapMirror @ 15 min

DR site (60 Ox)

Head 1

Head 2

Head 3

NetApp filer

NFS

Internal drives

IQSS Storage and Backup
Cloud backup options

• Local storage
  • Copies are made remotely on cloud storage for run-time
  • New or updated data is stored on local storage
  • High transfer costs?

• Local backup
  • Amazon S3 is redundantly backed up by default
  • Amazon EBS is not
  • Periodically back up local copies of data
    – Incremental backups?
    – Somewhat fixed transfer costs?
Cloud backup (Local storage)

User (applications) → Internal drives → NetApp filer

- HTTP
- SFTP
- NFS
- CIFS

IronKey (security Keys) → Internal drives

NetVault @ 1 day

NetApp filer → Tape drive

- HTTP
- SFTP
- NFS
- NDMP
- SnapMirror @ 15 min

Tape drive

DR site (60 Ox)

Head 1

Head 2

Head 3

HTTP

SFTP

Internal drives
Cloud backup (Local backup)

User (applications)

- HTTP
- SFTP
- NFS
- CIFS

IronKey
(security Keys)

Internal drives

Server

NetApp filer

- Head 1
- Head 2

NetVault
@ 1 day

@ 1 day (incremental)
@ 1 month (full)

SnapMirror
@ 15 min

Tape drive

DR site (60 Ox)
Questions

Backup
• We are considering adding a Virtual Tape Library (VTL) in place of the Tape Backup Device, i.e., *most* backups become disk and not tape backups. Thoughts?

Cloud
• For the cloud model, do we bother storing data locally? (other than confidential data)
• For the cloud model, what’s the difference between backing up locally and backing up at an alternative cloud vendor?
• What does disaster recovery mean to us in the cloud?
• Since storage for each application can be tied to the cloud image (AMI) and its associated storage, how does file service change in the cloud?