

# UW-IT

# Backups & Archives

Powerful, Flexible, Affordable

UW-IT TechTalk  
February 19, 2015

# Agenda

- Definitions
- Yesterday
- Today
- Tomorrow
- Your thoughts

# Backups Defined

- Data is “hot”
- Primary data copy is on first-tier storage
- Changes automatically preserved daily
- Bias toward small files
- Relatively expensive (many small files problem)

# Archives Defined

- Data is “cold”
- All/only copies reside in the Archives
- Additions/Modifications/Removals all manual
- Bias toward very large files
- Relatively inexpensive (manage few large objects)

# Backups + Archives

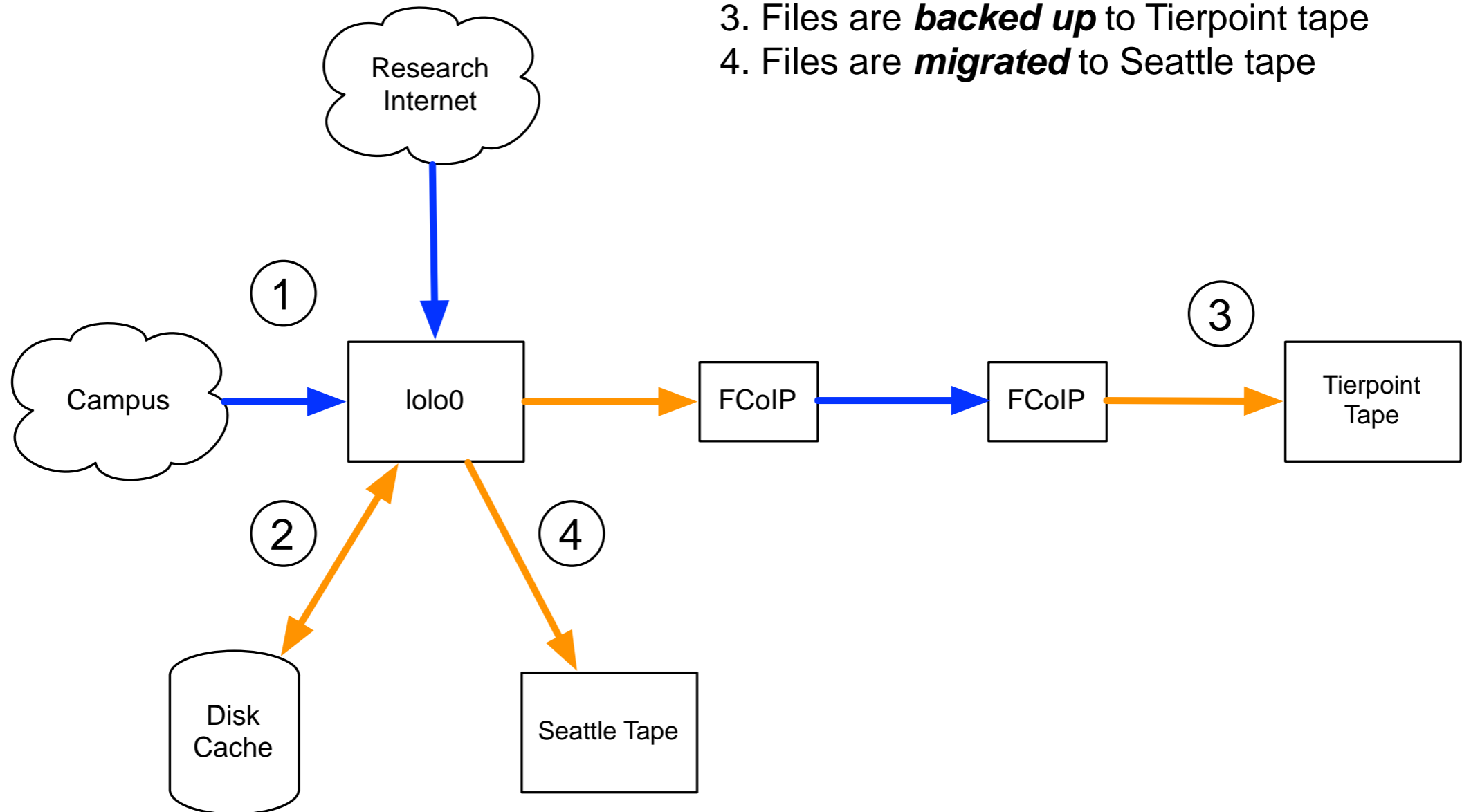
- Most (~90%?) of all data is “cold”
- Move cold data to Archives, reduce first tier storage demand by ~90%
- ~90% reduction in first tier storage = ~90% reduction in backups
- Potential for dramatic cost savings
- Potential for dramatic improvements in data management practices.

# Archives Today

- One node from the 4-node lolo cluster
- Between 2TB and 20TB each night
- Typically < 2,000 files each night
- 770TB subscribed, 370TB stored
- ~350 tapes in use across two sites
- 37 customers, Hyak by far the largest

# Archive Service Design

1. Customers upload blobs to lolo via ssh
2. Uploaded files are cached on 20TB disk
3. Files are **backed up** to Tierpoint tape
4. Files are **migrated** to Seattle tape



# Backups Yesterday

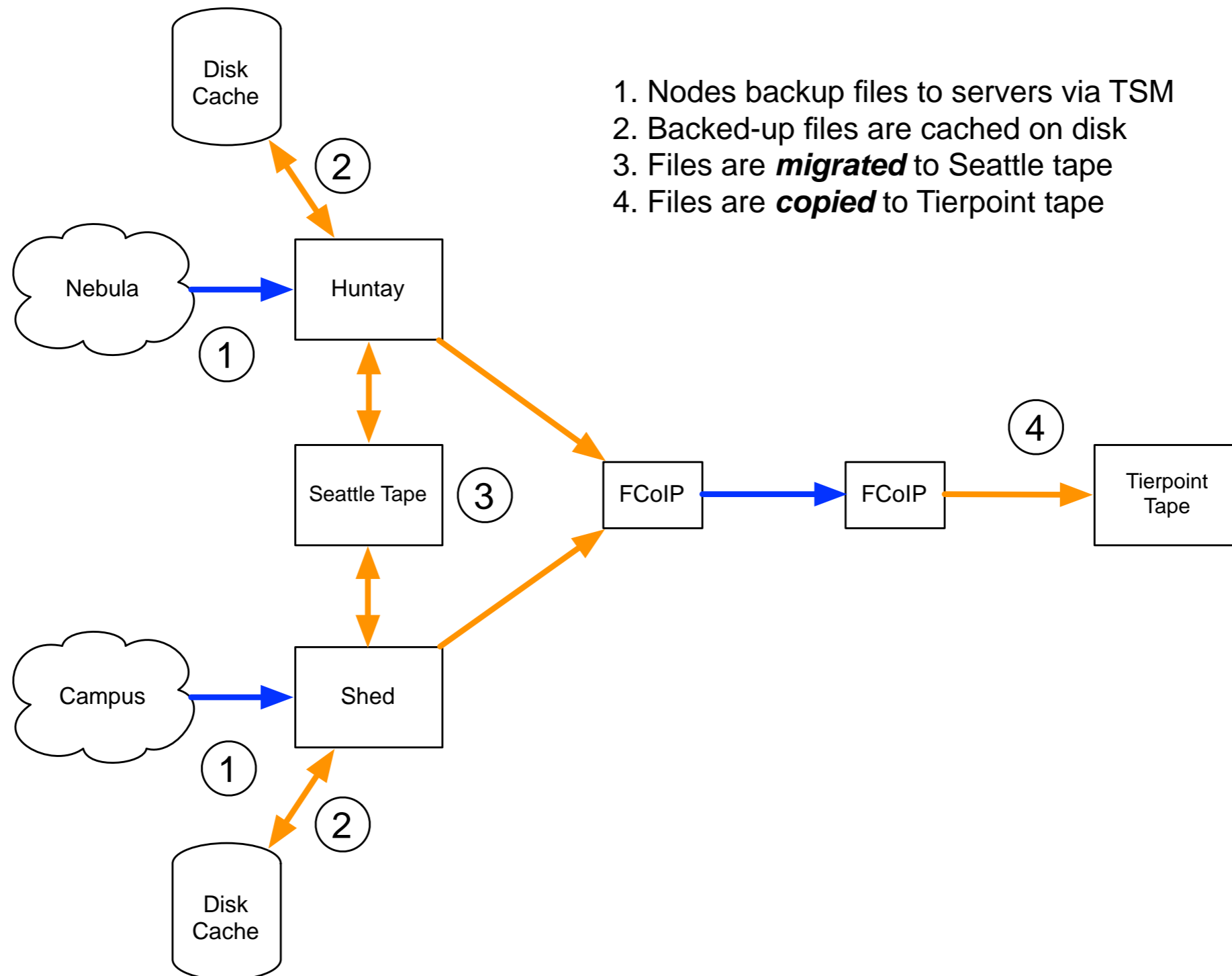
- > ?? nodes (lots, but difficult to determine #)
- > 5 hosts/servers (2 OSes/HW architectures, nonstandard)
- > 5 tape technologies
- > ?TB, >?M files each night (not great reporting)
- > 200TB, > 200M files stored at each site
- > 5,000 tapes in use across two sites



# Backups Today

- > 1,000 nodes (improved reporting)
- > 2 hosts/servers (2 OSes/HW architectures, nonstandard)
- 1 tape technology (**5 x reduction**)
- > 10TB, >2M files each night (improved reporting)
- > 500TB, > 500M files stored at each site (**2 x increase**)
- > 500 tapes in use across two sites (**10 x reduction**)

# Backups Today Design



1. Nodes backup files to servers via TSM
2. Backed-up files are cached on disk
3. Files are **migrated** to Seattle tape
4. Files are **copied** to Tierpoint tape

# Backups Tomorrow: Goals

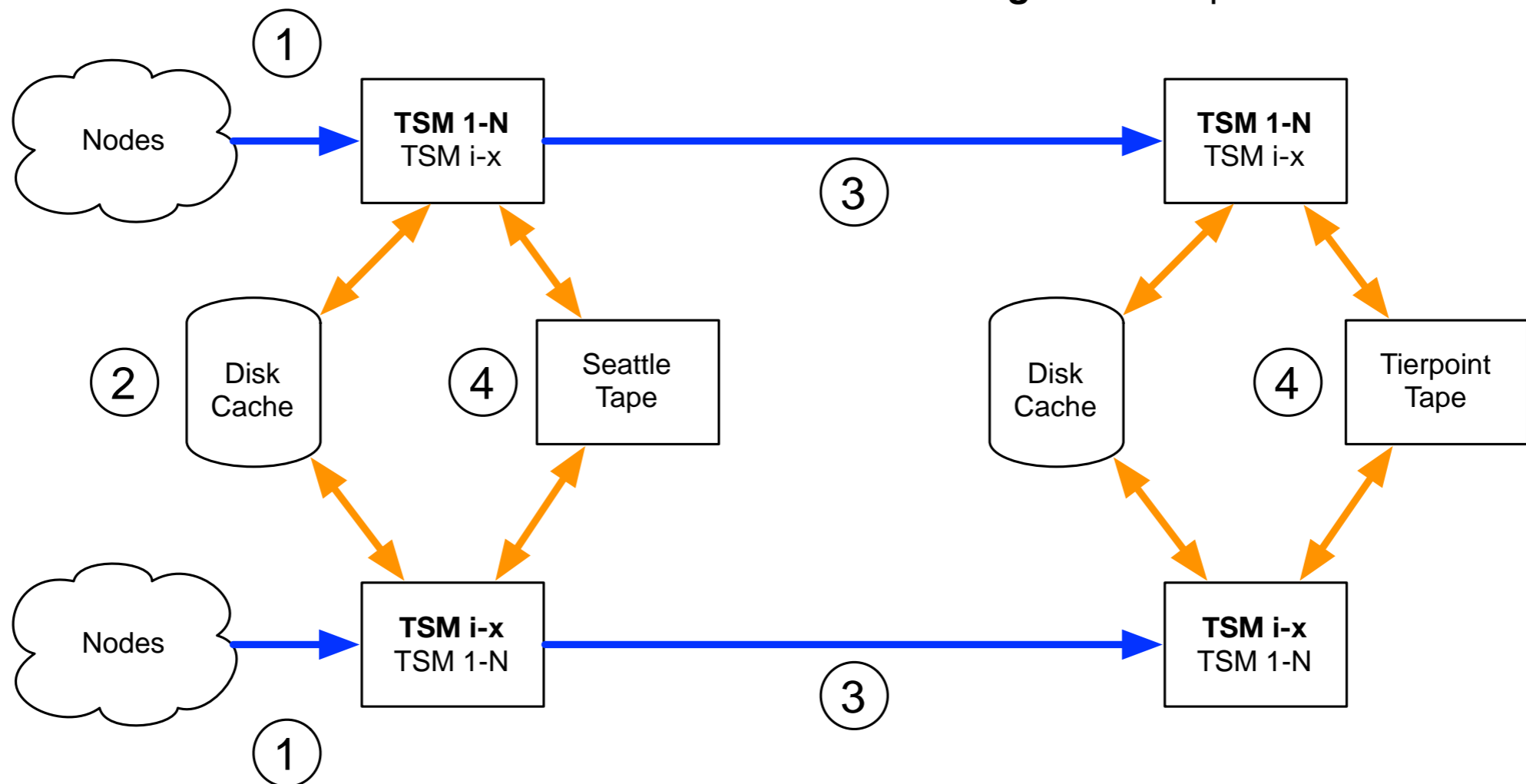
- Bring the service into the enterprise design fold
  - 1 x OS, 1 x HW Architecture, All standard
  - Standard toolset, layered design
- Reduce systems management effort
- Improve resilience and performance
- Improve scalability

# Backups Tomorrow: Features

- High Availability
- Improved GR/DR
- Prepared for Archive integration
- Prepared for enhanced features

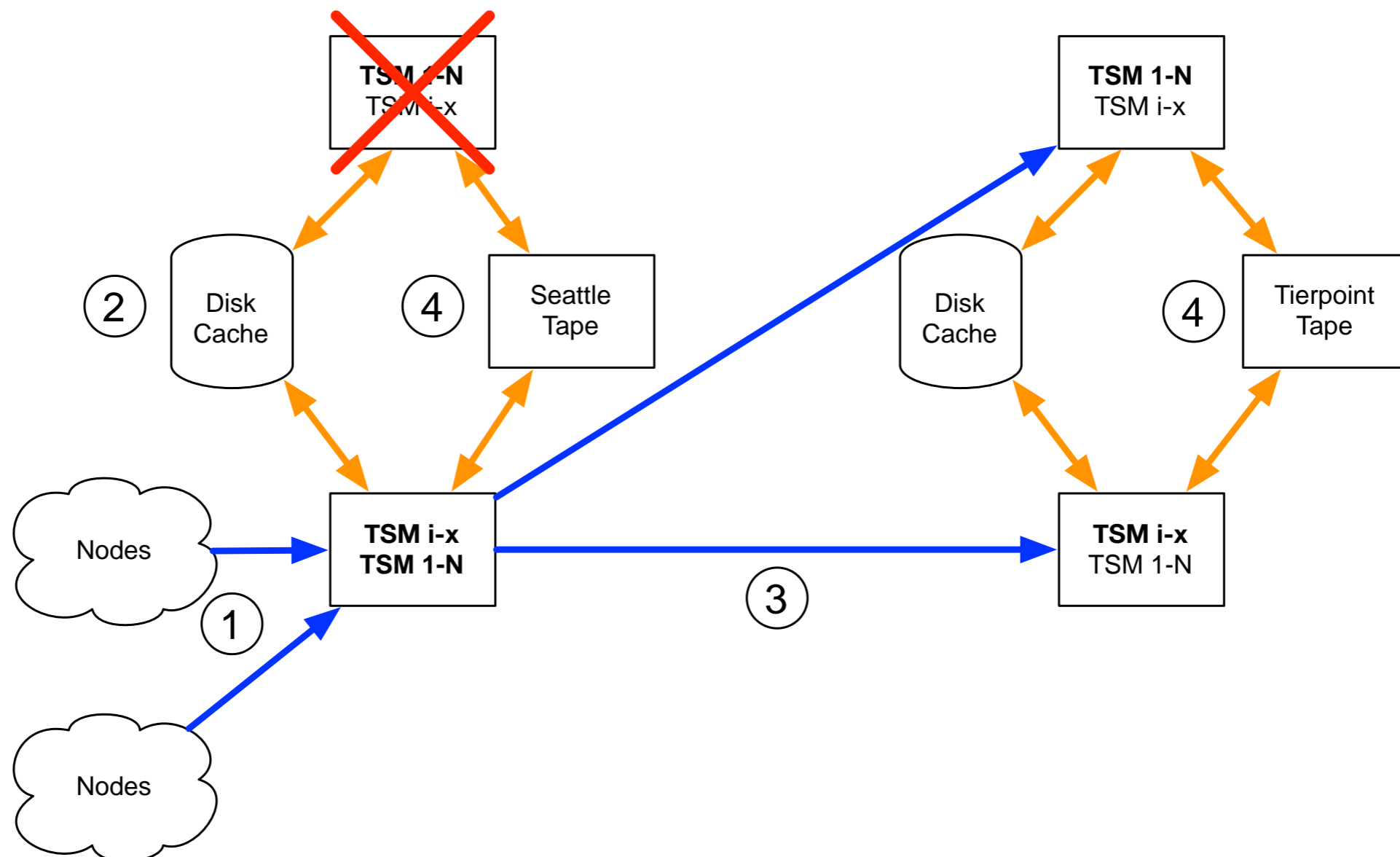
# Backups Tomorrow Design

1. Nodes backup files to servers via TSM
2. Backed-up files are cached on disk
3. Files are **replicated** to Tierpoint servers
4. Files are **migrated** to tape



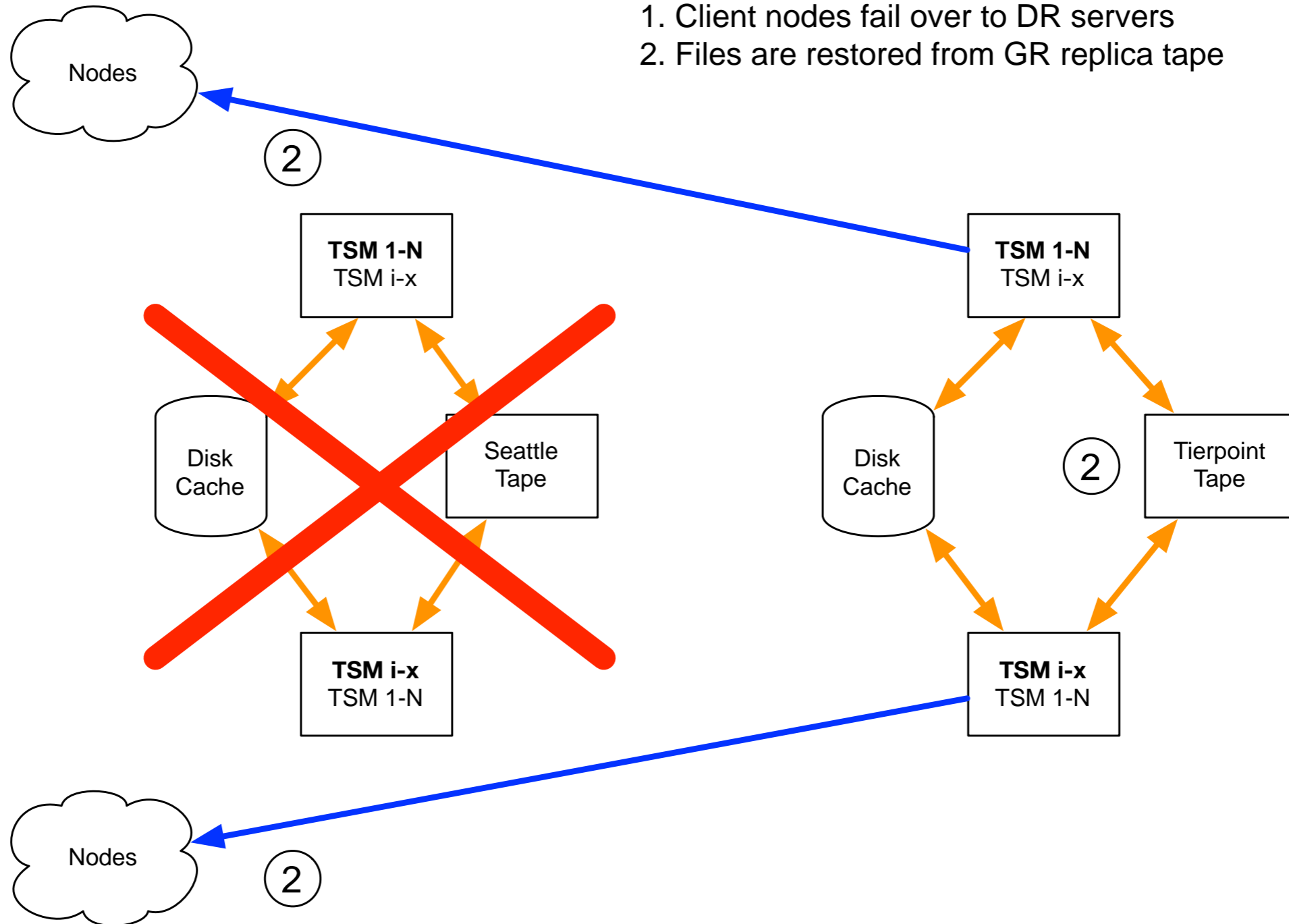
# Backups Tomorrow HA

1. Nodes backup files to servers via TSM
2. Backed-up files are cached on disk
3. Files are **replicated** to Tierpoint servers
4. Files are **migrated** to tape



# Backups Tomorrow **GR/DR**

0. Backup operations are suspended
1. Client nodes fail over to DR servers
2. Files are restored from GR replica tape



# Your Thoughts?

- Better support for VMWARE, databases, etc.?
- What sort of data would you like to see?
  - Details about your node's backups?
  - Details about the overall service?
- Wish List?



# Service Catalog

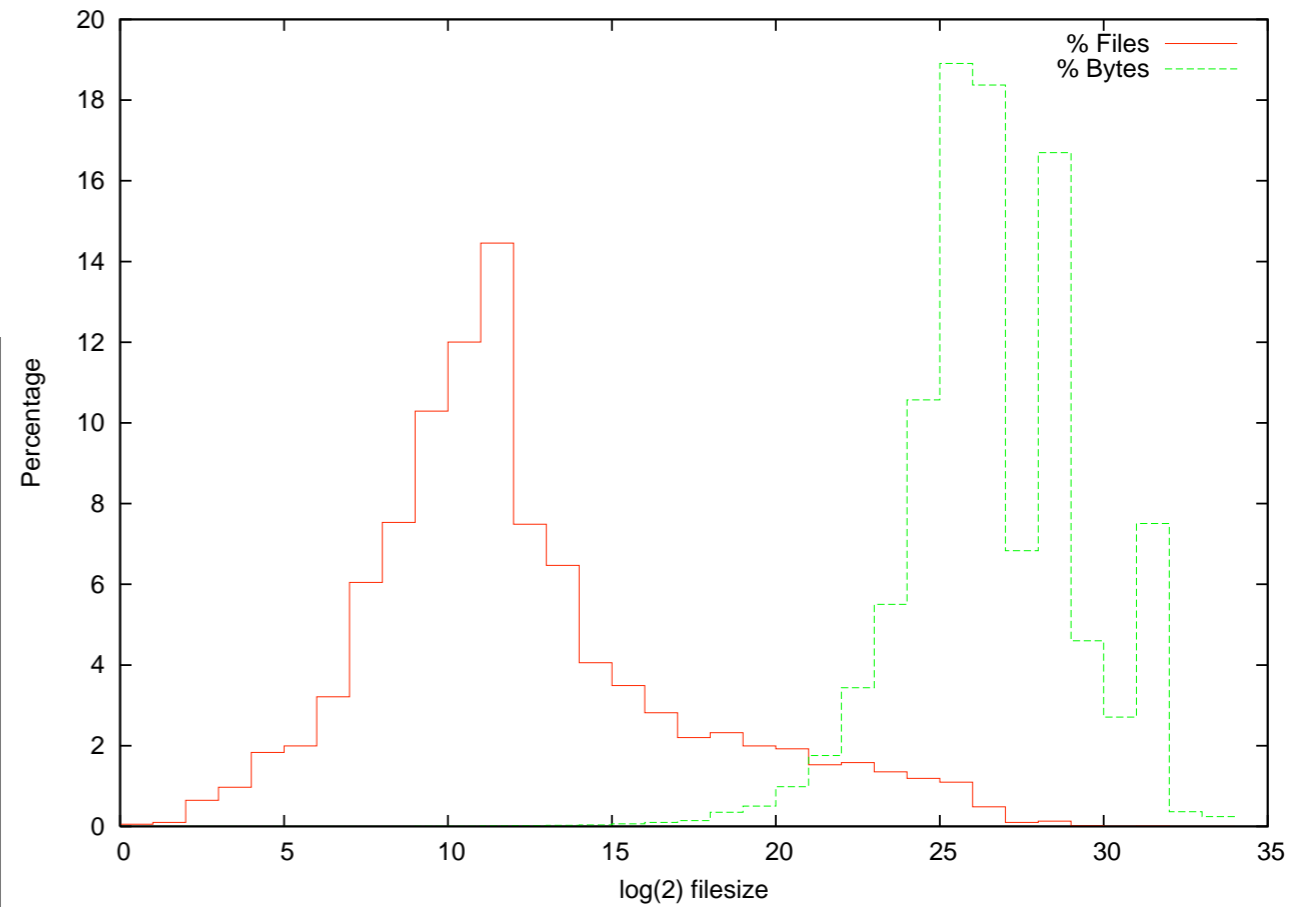
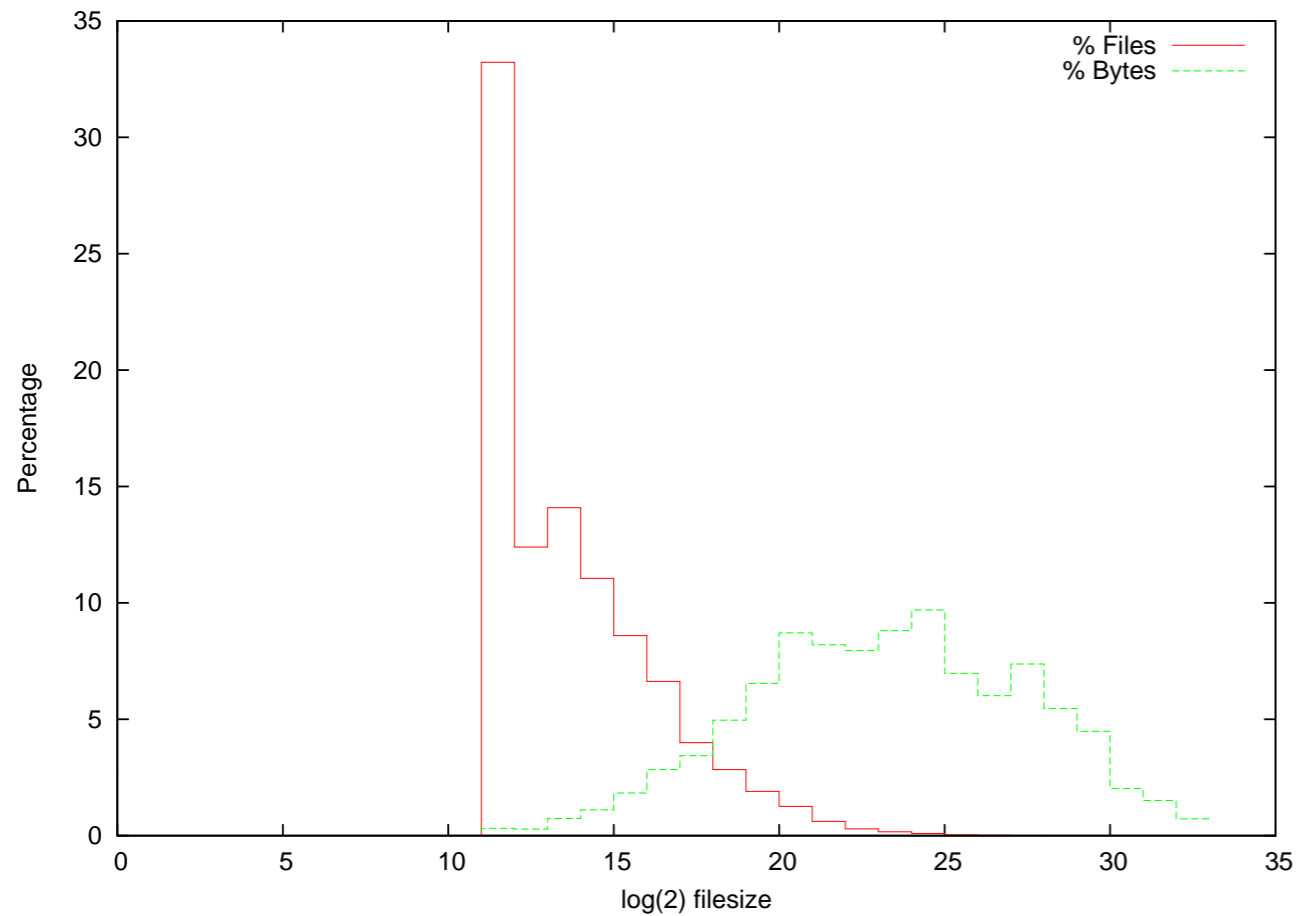
- Iolo Archives
  - <http://depts.washington.edu/uwtscat/archivestorage>
- Backup Service
  - <http://depts.washington.edu/uwtscat/databackuparchive>

# Supplemental Slides

- File size and bytes distributions
- Transfer rates for Iolo Archives
- Prices

# File Size Distribution

Nebula in 2008



An Astrophysics Group  
in 2006

# Iolo Archive Recall Rates

| MB        | Tape Load/Seek<br>(Sec) | Tape Read<br>(MBs) | Recall Time<br>(sec) | Recall Speed<br>(MBs) |
|-----------|-------------------------|--------------------|----------------------|-----------------------|
| 1         | 90                      | 125                | 90.01                | 0.01                  |
| 10        | 90                      | 125                | 90.08                | 0.11                  |
| 100       | 90                      | 125                | 90.80                | 1.10                  |
| 1,000     | 90                      | 125                | 98.00                | 10.20                 |
| 10,000    | 90                      | 125                | 170.00               | 58.82                 |
| 100,000   | 90                      | 125                | 890.00               | 112.36                |
| 1,000,000 | 90                      | 125                | 8,090.00             | 123.61                |

# Some Prices

|              | ¢/GB/Month | \$/TB/Year | @50% Use |
|--------------|------------|------------|----------|
| UW-IT Backup | 6.0        | 720        | 720      |
| Iolo Archive | 0.9        | 103        | 206      |
| AWS Glacier  | 1.0        | 120        | 120      |