GPFS with underlying ZFS block devices

Christopher Hoffman

June 10th

LA-UR-16-23957

Overview

- Motivation
- General Configuration
- Use Cases
  - Archive
  - Campaign Storage
  - MarFS

LA-UR-16-23957

Motivation

- Cheaper components
- Added functionality
  - Compression
  - Extra Parity
  - Copy-on-write
- Takes out one unknown component
- In-house expertise
  - Over 40PB of ZFS deployed at LANL

LA-UR-16-23957

General Configuration

- ZFS options
  - recordsize= 1M
  - sync=always
  - compression=lz4
- zvols
- nsddevices
  - echo "zdX generic”
- 2x data replication

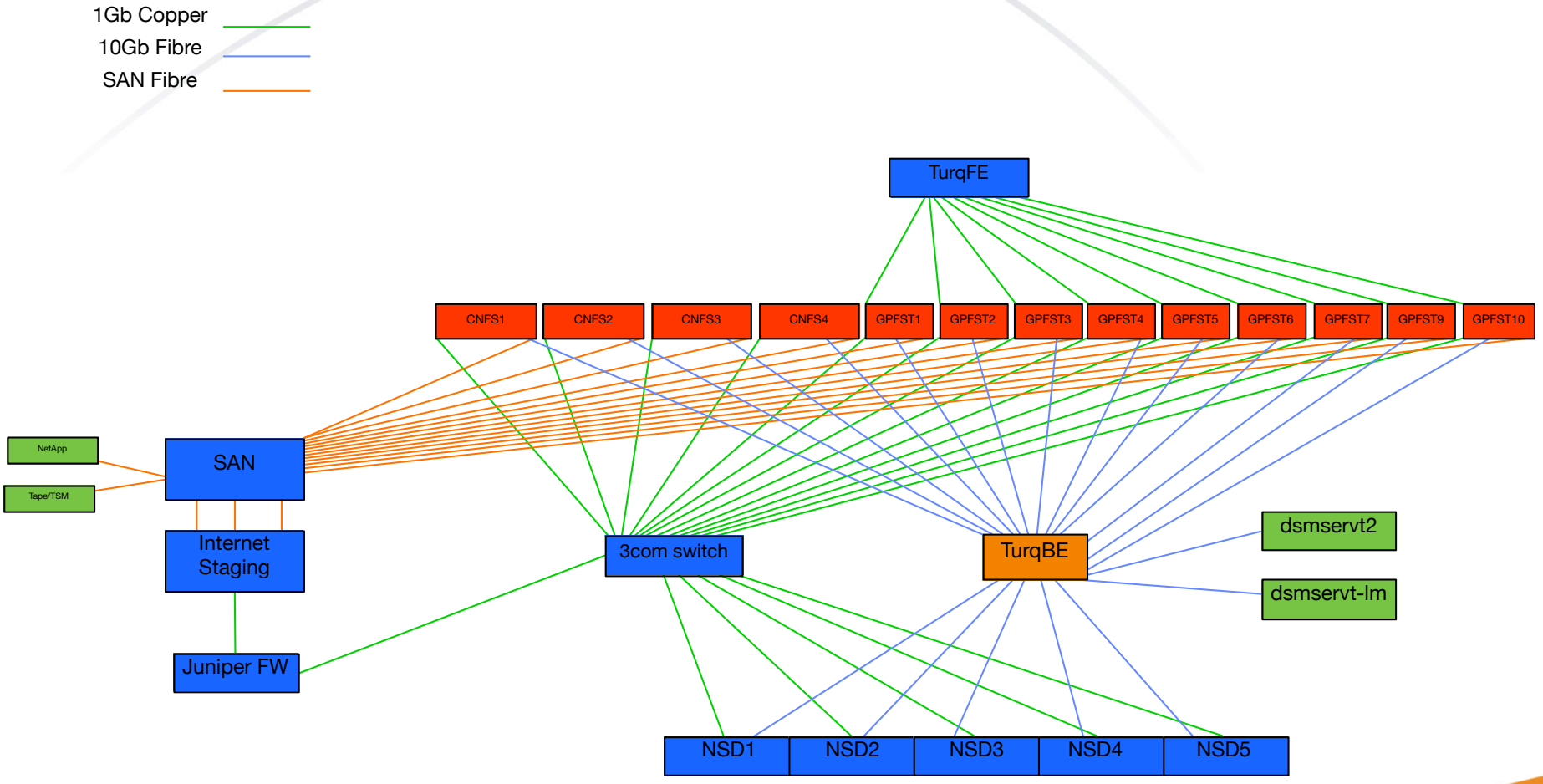
LA-UR-16-23957

Archive Overview

- Long term storage in Institutional Computing
- Old Version:
  - NetApp for metadata & landing zone
  - Small file copy on disk (<1GB)
    - RAID6
  - Copy of small files on tape
  - Single copy of large files on tape
  - HSM

LA-UR-16-23957

Old Archive



LA-UR-16-23957

UNCLASSIFIED

Archive Overview (cont'd)

- New Version:
  - NetApp for metadata & landing zone
  - Premigrate all files to tape
  - Copy files to new ZFS Disk Pool

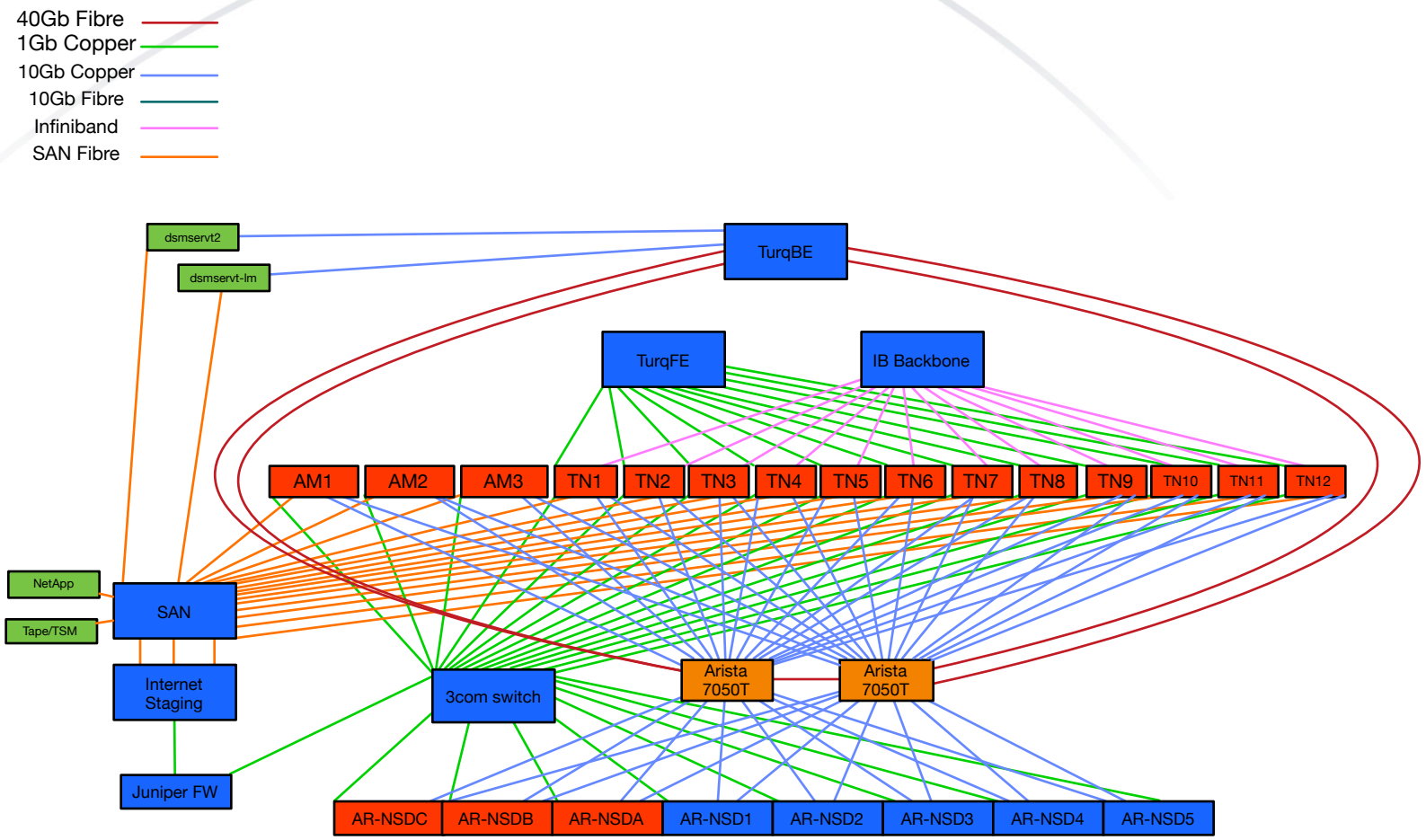
LA-UR-16-23957

Archive Overview (cont'd)

- 8x GPFS NSD Servers (TOSS)
  - Dual bonded 10Gb Copper Ethernet
  - 4x JBOD per server
    - 60x 5TB SMR Seagate via 6Gb SATA
    - 3x RAIDZ3 16+3
  - NetApp e5560
- MLAG
- 8GB/s
- 1.24x compression

LA-UR-16-23957

New Archive



LA-UR-16-23957

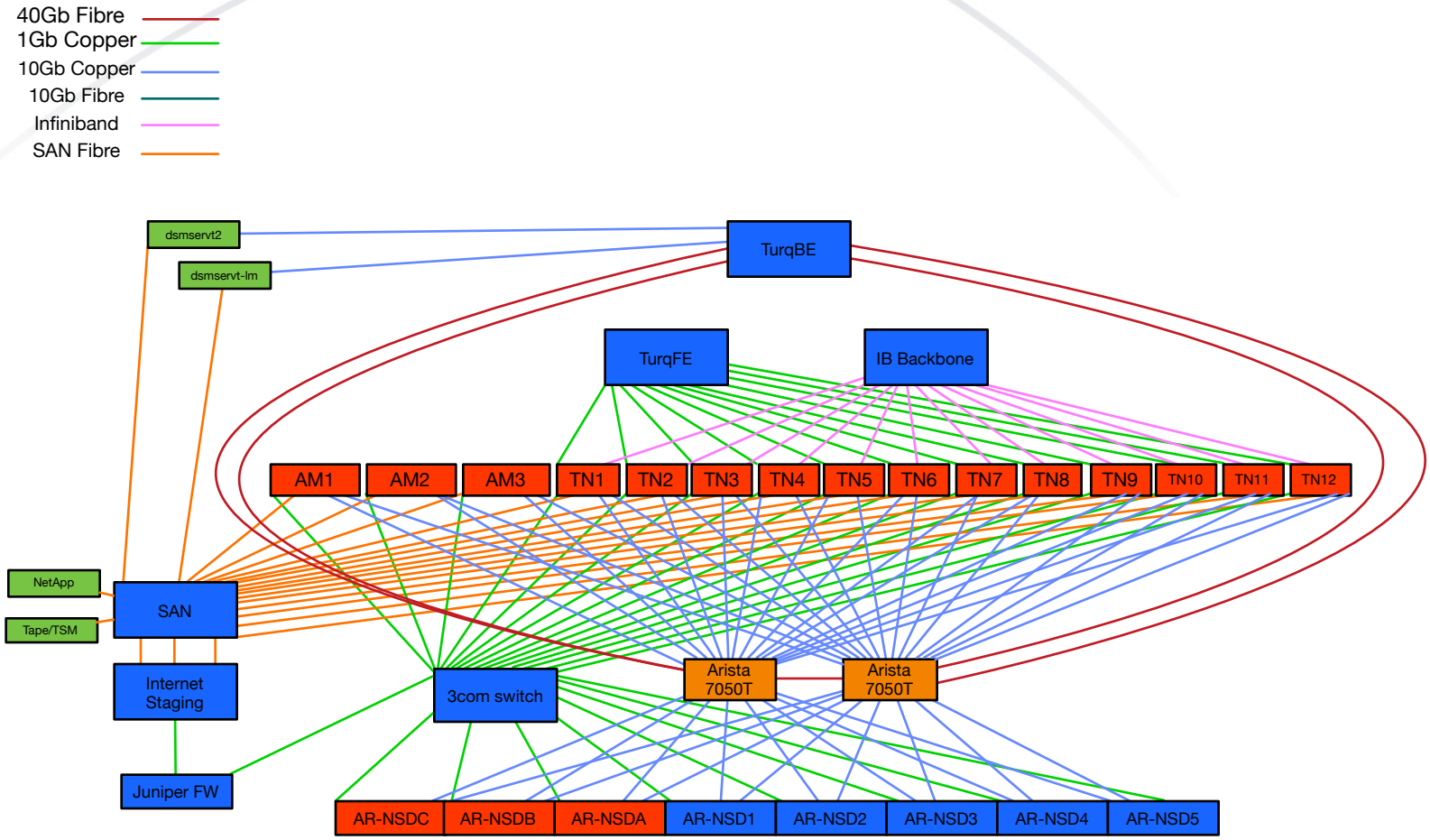
UNCLASSIFIED

Campaign Storage Configuration

- Added to existing GPFS NSD Servers
- 8 GPFS NSD Servers (TOSS)
  - 84x 4TB PMR Seagate Drives
    - dataOnly
  - 2x 100GB SSD
    - metadataOnly

LA-UR-16-23957

Campaign Storage



LA-UR-16-23957

UNCLASSIFIED

MarFS Overview

- Next Generation LANL archive
- Writes data to Object Store
- Writes metadata to GPFS Cluster with SSDs
  - via RDMA
- Limited fuse interface
  - ro for data ops
  - rw for metadata ops
- Pftool for data movement

LA-UR-16-23957

MarFS/GPFS Integration

- Uses fast inode scans frequently
- Makes use of libgpfs:
  - Garbage collection
  - Packing
  - Quotas

LA-UR-16-23957

Questions?

- MarFS: http://github.com/marfs

LA-UR-16-23957