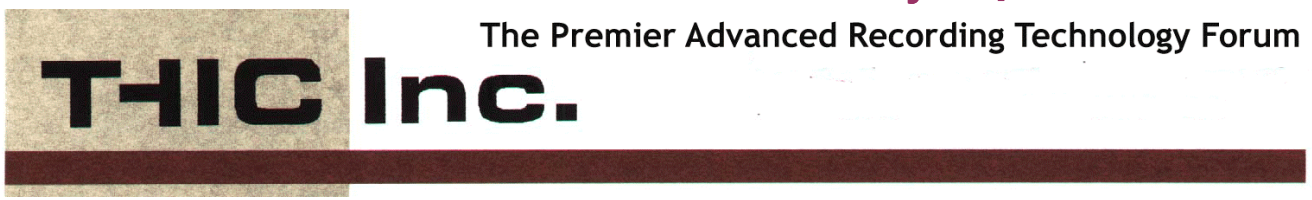
High I/O Access Rates to the HPSS Archive at SDSC

Tom Sherwin
San Diego Supercomputer Center
10100 Hopkins Dr., San Diego CA 92093-0505
Phone:+01-858-534-5110 FAX: +01-858-534-5152
E-mail: sherwint@sdsc.edu

Presented at the THIC Meeting at the
Bahia Hotel
San Diego CA
on January 16, 2001



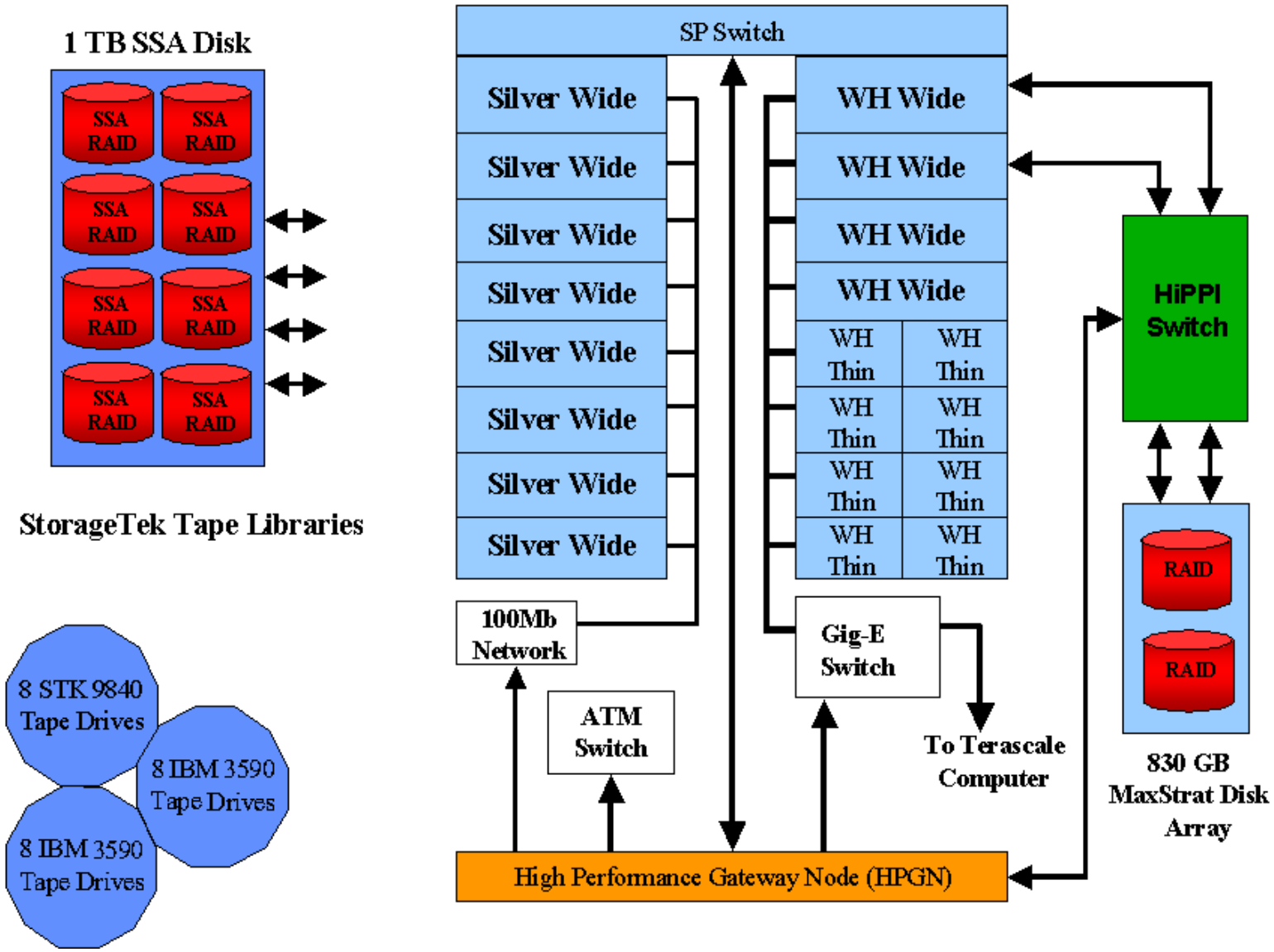
What is SDSC?

- One of three NSF funded centers
  - SDSC
  - NCSA
  - Pittsburgh
- Exist to provide unclassified NSF funded research to be done on state-of-the-art resources.

The HPSS System

- **H**igh **P**erformance **S**torage **S**ystem
  - http://www4.clearlake.ibm.com/hpss
- IBM led collaboration between government and industry
- Approximately 20 installed sites worldwide
- SDSC is currently the largest HPSS system with more than 220 Terabytes stored in 14 million files

HPSS Configuration



Compute Platforms

- Predecessor systems
  - Cray T90 – 14 CPU, 24 Gflops
  - Cray T3E – 256 CPU, 154 Gflops
  - Tera MTA – 8 CPU, 8 Gflops
- Blue Horizon
  - IBM SP with 144 Nighthawk II nodes
  - 8 CPU and 4 GB memory/node
  - 1.7 Teraflops peak!
  - 5 Terabyte GPFS filesystem

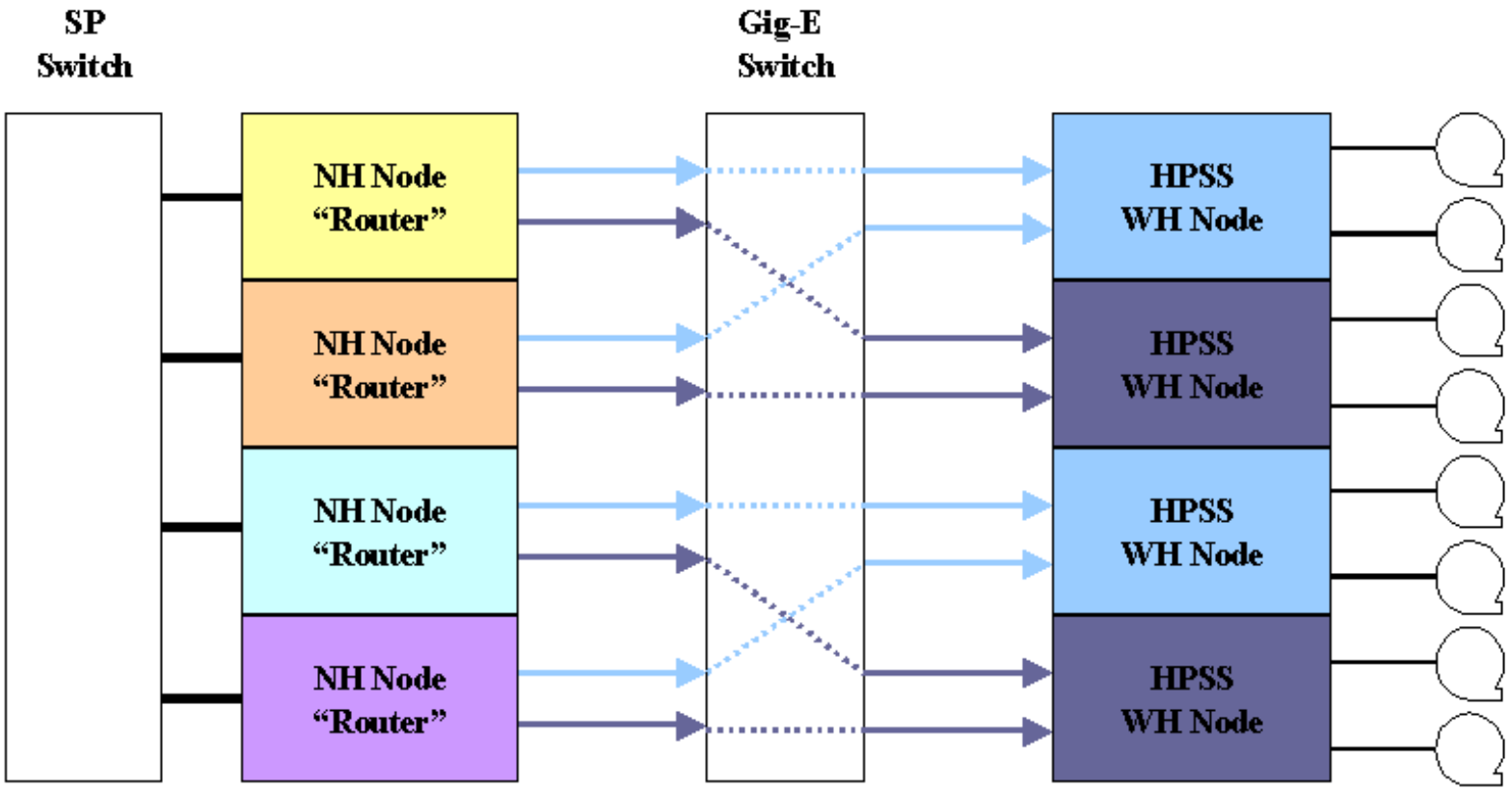
The Challenge

- Show that data can be moved at one Gigabit/sec (120 MB/sec) directly to tape
- Simple concept
- Harder to implement

Assumptions and Limitations

- Tape drives are 15MB/sec max. 8 are required for the desired rate
- Gigabit adapter != 1Gb/sec
- IBM SP Switch limited to 122MB/sec

Networking Concept



Surprises

- Nodes incorrectly configured or inadequate for 3590 performance
- HPSS client software required changes to implement read-ahead buffer pool
- HPSS does not like having its resources monopolized for long periods of time

Results

Transfer Type	Sparse File	Uncompressible File	Scientific Data
One one-way	16.4 MB/sec	11.4 MB/sec	16.3 MB/sec
One two-way	29 MB/sec	23.7 MB/sec	25.5 MB/sec
Two two-way	52.1MB/sec	45.5 MB/sec	50.4 MB/sec
Four two-way	108 MB/sec	89.6 MB/sec	106.3 MB/sec
One eight-way	36.6 MB/sec	30.8 MB/sec	31.7 MB/sec

Future Work

- Recent SP switch upgrades should allow better GPFS performance
- Need to see how jumbo (9000 byte) MTU effects results
- New fiberchannel SAN attached disk and disk striping

Conclusions

- High I/O rates can be achieved using striped devices
- Tuning is not limited to the network
- Test the entire configuration, not just the pieces