Billions of Collisions: Petabytes of Data

Large physics experiments have millions of data files on tens of thousands of tapes.

Storage systems must manage data placement and retrieval to the limits of practicality and affordability.

Data from tape must be readily accessible to hundreds of Physicists for analysis both locally and over the global network.

High-Level Storage Requirements in HEP

Storage systems (disk & tape) are a major fraction of computing costs.

- Dominant tape access is reading
- Granular datasets, with many files
- Files read in parallel and independently
- Reading data optimized by clustering files on tape
- Data access software benefits from seeing the map of files on tape
- Small amount of data loss traded for much larger capacities
- Normally only one copy of irreproducible data

Large Scale Media Use

Fermilab Mammoth II (60 GB)
- 17,000 tapes per year for Run II data
- Reducing media cost per petabyte for data storage an important factor
- Collaborating institutions support same commodity media
- Inexpensive commodity drives allow for future hardware upgrades
- Improved quality of commodity drives

SLAC
STK 9840 (20GB) now
STK 9940 (60GB) soon
- 5,000 tapes/year for BaBar data
- All BaBar tapes in robotic silos
- Collaborating institutions exchange data over network

High Performance Storage System - HPSS

Fermilab Supports 16 groups & experiments
- 8mm tape import/export facilities
- Fermilab provided user interface wrapper
- 23 TB stored (290,000 files)
  (30 TB total storage)
- Transfers average 250 GB/day

SLAC
Principal storage system for scientific data
Totals (Aug 2000):
- 174 TB
- 98,000 files
  (3 GB mean file size)

Activity per Month:
- Add 30 TB
- Read/write 13,280 files

Enstore Mass Storage System at Fermilab

Storage systems for High Energy Physics transfer immense amounts of data to and from tape continuously. The DZero experiment needs 150 MB/sec of continuous transfer rate, and the ALICE experiment at CERN, 1000 MB/sec.

System architectures thus must emphasize scalable and parallel reading and writing of tape carried out by loosely coupled, independent computers using third party transfers.

The Fermilab Enstore system, deployed for DZero, illustrates this type of system. New Mover computers use commodity "white box" Intel server computers, with connections to a well designed commodity network. Each Mover computer transfers 20 MB/sec between the network and a pair of Mammoth II tape drives. Commodity computers, networks and tape drives set the cost of a 20 MB/sec increment of throughput around $11,000.

The active display illustrates the overall dataflow. A data mover is paired with file transfer requests. An automated tape library mounts the relevant tape, then data is transferred between tape and the network. The system satisfies requests, possibly from various computers, for access to the same tape, then the tape is put away.

Billions of Collisions: Petabytes of Data

Large physics experiments have millions of data files on tens of thousands of tapes.

Storage systems must manage data placement and retrieval to the limits of practicality and affordability.

Data from tape must be readily accessible to hundreds of Physicists for analysis both locally and over the global network.

High-Level Storage Requirements in HEP

Storage systems (disk & tape) are a major fraction of computing costs.

- Dominant tape access is reading
- Granular datasets, with many files
- Files read in parallel and independently
- Reading data optimized by clustering files on tape
- Data access software benefits from seeing the map of files on tape
- Small amount of data loss traded for much larger capacities
- Normally only one copy of irreproducible data

Large Scale Media Use

Fermilab Mammoth II (60 GB)
- 17,000 tapes per year for Run II data
- Reducing media cost per petabyte for data storage an important factor
- Collaborating institutions support same commodity media
- Inexpensive commodity drives allow for future hardware upgrades
- Improved quality of commodity drives

SLAC
STK 9840 (20GB) now
STK 9940 (60GB) soon
- 5,000 tapes/year for BaBar data
- All BaBar tapes in robotic silos
- Collaborating institutions exchange data over network

High Performance Storage System - HPSS

Fermilab Supports 16 groups & experiments
- 8mm tape import/export facilities
- Fermilab provided user interface wrapper
- 23 TB stored (290,000 files)
  (30 TB total storage)
- Transfers average 250 GB/day

SLAC
Principal storage system for scientific data
Totals (Aug 2000):
- 174 TB
- 98,000 files
  (3 GB mean file size)

Activity per Month:
- Add 30 TB
- Read/write 13,280 files

Enstore Mass Storage System at Fermilab

Storage systems for High Energy Physics transfer immense amounts of data to and from tape continuously. The DZero experiment needs 150 MB/sec of continuous transfer rate, and the ALICE experiment at CERN, 1000 MB/sec.

System architectures thus must emphasize scalable and parallel reading and writing of tape carried out by loosely coupled, independent computers using third party transfers.

The Fermilab Enstore system, deployed for DZero, illustrates this type of system. New Mover computers use commodity "white box" Intel server computers, with connections to a well designed commodity network. Each Mover computer transfers 20 MB/sec between the network and a pair of Mammoth II tape drives. Commodity computers, networks and tape drives set the cost of a 20 MB/sec increment of throughput around $11,000.

The active display illustrates the overall dataflow. A data mover is paired with file transfer requests. An automated tape library mounts the relevant tape, then data is transferred between tape and the network. The system satisfies requests, possibly from various computers, for access to the same tape, then the tape is put away.