

Evolution of Storage Software

T. M. Ravi
Mimosa, Inc.

Phone:408-464-7075

E-mail: ravitm@pacbell.net

**Presented at the THIC Meeting at the Sony Auditorium,
3300 Zanker Rd, San Jose CA 95134-1940**

March 4-5, 2003

The Big Three: Backup

- Periodic copy of data (or changed data) to tape
- Tape is inexpensive, easy to transport and stable over time
- Data restore is broken
 - Process is cumbersome
 - Restore times are high
 - Requires administrator intervention and is manual
- Lack of confidence in backup reliability – backup process, drives and media are error prone
- Imposes heavy performance load on enterprise server
- Open files not backed up
- Does not address Business Continuity needs
 - Restoring data from nightly backup is too slow
 - Restored data could be at least a day old

The Big Three: Replication

- Synchronous mirroring ensures data availability in the event of catastrophic failure
 - Immediate writes to remote mirror
 - Write not acknowledged till data is committed to mirror also
- Very expensive
 - Write latency penalty requiring dedicated high speed line
 - High network bandwidth costs
 - Cost of enterprise storage is doubled
- Recovery from disconnection of mirrors is complex
- Accidental deletion is instantly propagated - still need backup to restore data that has been deleted or corrupted

The Big Three: Archive and HSM

- Manual or automated migration of older data to tape
- Extend enterprise storage
- Improve enterprise server performance
- Time and effort it takes to locate and access archived data discourages data access
- Manual archiving is too labor intensive
- Limited market success due to mismatch between tape and disk and high costs of libraries
- Need for data life cycle management - New regulations and policies for data and email retention

The Newcomers: Snapshot

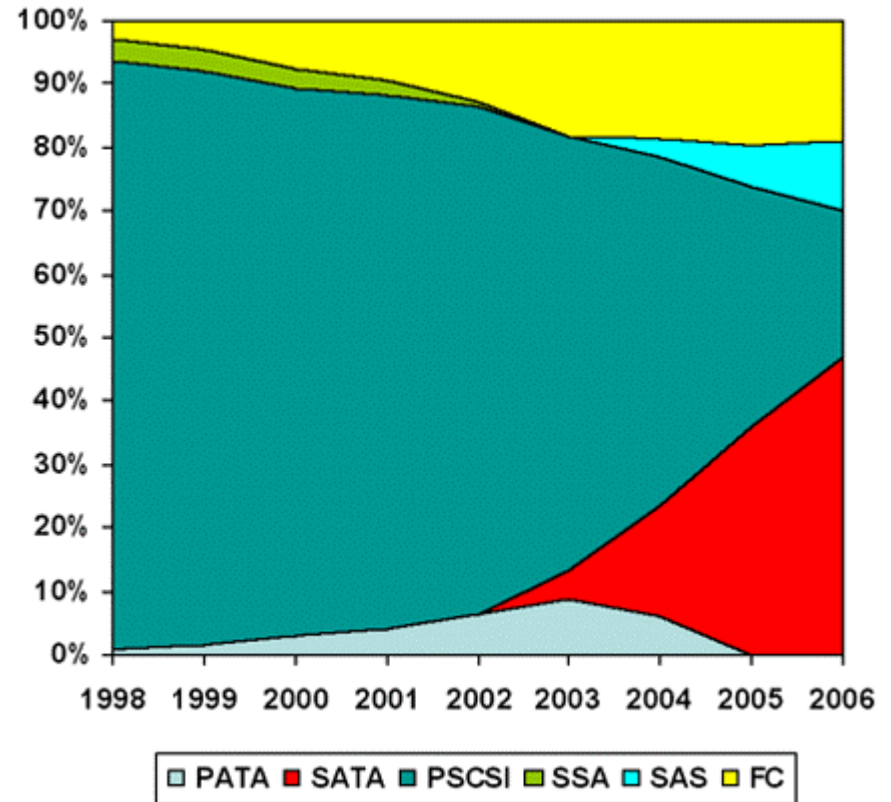
- Consistent point-in-time view of data
- Emerging as key enabling technology for data protection of applications that are live 24 x 7 – addresses the open file backup challenge
- Process
 - Freeze application (Quiesce)
 - Flush buffers
 - Take Snapshot image
 - Thaw
- Provided by disk arrays and NAS
- Soon to be supported by OS
 - Microsoft .NET VSS

The Newcomers: Asynchronous Replication

- Mirrors are periodically synchronized
 - Removes latency overhead for writes
 - Use shared low cost lines
- Snapshots (or changes) are copied to the mirror
- Mirror lags behind primary storage
- Update frequency can be tuned
 - Data loss window vs. network bandwidth cost and server load
- Mirror storage need not be the same as primary storage

Enterprise Disk Drive Market

- Emergence of new class of low cost disk storage – Serial ATA
- Transfer rates of 150 (now), 300 & 600 MB/sec
- Cost driven down by desktop volume
- SATA disk costs catch up with tape in 2002 (IDC)
- SATA array costs catch up with tape libraries in 2006 (Gartner)



(Source: IDC, August 2002)

Evolution of Data Protection

- New class of data protection software that blurs the distinction between Disaster Recovery, Backup and Archive
- Persistent storage of snapshots – enable disaster recovery
- Store multiple snapshots – enable restore of lost file or data object
- Store in vault server with low-cost SATA or SCSI disks
 - Traditional direct access file or database interface
 - Secure - Integrate with access control
 - Robust - With RAID
- Local and remote vaults
- Where does tape fit?
 - Periodic tape backup of vault server
 - Long term archival

Benefits

- Instantaneous recovery of file systems, databases etc. to a previous point-in-time in case of disaster
- Ability for users to restore their own deleted, lost or corrupted files – backup/restore from disk
- Archive of older data and consolidation of storage
- Secure repository for reference information
- Platform for decision support applications
 - Applications such as data mining do not stress online application server

Conclusion

- Customer needs driving new cost effective disaster recovery solutions
- Significant evolution of Backup
 - Reduced complexity of backup
 - Tape emulation by disk
 - Disk-to-disk backup
- New generation of archive solutions
- New class of reference information servers
- Balanced data protection strategy that addresses cost, ease of use, reliability and need for data access throughout the life cycle

.... low cost disk (e.g. Serial ATA) is the enabler