

# Back to the Future

Gerhard Schneider  
Rechenzentrum Uni Freiburg  
gerhard.schneider@rz.uni-freiburg.de



## Are tapes still useful?

- Disk v. Tape over the years:

Year	Tape	Disk	Tape per GB	Disk per GB
1992	800MB 3480	1GB SCSI/ € 1400		€ 1400,--
1998	20GB 3480/ €60	9GB SCSI/ € 500	€ 3,--	€ 56,--
2001		100GB IDE/ € 400		€ 4,--
2003	40GB Magstar €65	250GB IDE/ € 350	€ 1,65	€ 1,40

## Cost of Backup



- Already mentioned in my 2001 talk:  
Tapes require an infrastructure
  - Tape robot, tape drives, maintenance contract
- So the cost per GB is even higher
- Example: DELL LTO-Library 136T
  - 3 LTO drives, 72 tapes, 7.2 TB capacity, 2 power supplies, SCSI: € 35.000
  - Cost of overhead: 5€ per GB
  - Do larger systems offer a better performance?? ☹

## Cost of Backup



- Cheap IDE-Raids are readily available:
  - Internal IDE
  - External SCSI/FC
- 4 TB cost € 10.000,--
- Backup of such RAIDs to tape is more expensive than the RAID
- Users have such devices!
  - Backup of this RAID to the said LTO library....
    - 2 copies to tape
  - costs around € 40.000,--



## Other Approaches to Backup



- So why not backup such a RAID to another RAID? (or two?)
  - Advantages: instant availability
    - Ever tried to restore 4TB from tape?
- Freiburg setup:
  - Fileserver mirrors its disk to a second RAID which is offsite (rsync ☺)
  - Second RAID is also available to users R/O
    - So maybe the just finished but deleted thesis is still living on the second RAID
    - If Fileserver crashes over the weekend, users have access to their data

## Consequences



- First step: since our robot has no more free capacity, we buy RAIDs instead of tapes
  - TSM can handle the resulting „chaos“ of media
- So why not throw out the tape robot altogether and use RAIDs for TSM?
  - The yearly maintenance charge buys 2 RAIDs
  - The robot is loaded with 1500 10GB-tapes
    - Warning: replacing just the tapes is expensive too!!
- The University of Tübingen has just done this
  - 100TB of disk space for backup only!

## Back to the Future!



- The moment we tried to sell the robot, IBM made a new announcement:

- New Magstar drives
- New tapes: 300GB for € 100 thus **€0,33 per GB**
- Upgrading the robot will be
  - 8 new drives = € 100.000,-- (??)
  - 1500 tapes = € 150.000,--
  - Resulting in 500 TB



- **Thus:**  
**€0,50 per GB**

## Back to the Future?!



- So with the 1TB tape on the horizon, our robot may continue to live for a few years.
- Message to the manufacturer:  
**Make tapes even cheaper – otherwise...**
- Message to the computer center  
*Whatever you do with tapes – you can do it with disks*

## Disks for Backup



- Reliability of IDE disks seems adequate
  - They die, but *RAID = redundant array of inexpensive disks* does handle the situation
    - just keep enough spare disks around
- Energy consumption is noticeable
  - 100 TB for backup = 25 RAID systems
  - This requires approx  $25 \times 400W = 10kW$ 
    - Maybe a bit less, but disks are spinning even when not needed

## Disks for Backup



- So why not have a rack of disks (all on one S-ATA bus) which are only turned on when needed?
  - How often can you switch on a disk?
  - How often may you mount a tape???
  - If a disk really breaks – is this due to a head crash or rather to a mechanical or controller failure?
    - The latter can be repaired – perhaps costly, but it can be repaired without loss of data
- No robot required: TSM could turn the disks on (only one per S-ATA bus) and write right away



## TSM's weak point: archive



- Computer center's business is not only backup – at least not any more
  - Long term availability of raw data
    - (guide of conduct in science: 10 years required / fraud)
  - Long term availability of precious data
  - Long term availability of living data
- TSM only provides a data sink
  - Who will remember the data structures?
  - Who will own the data in 50 years?
    - Not the node to which data is linked today

## TSM and archive



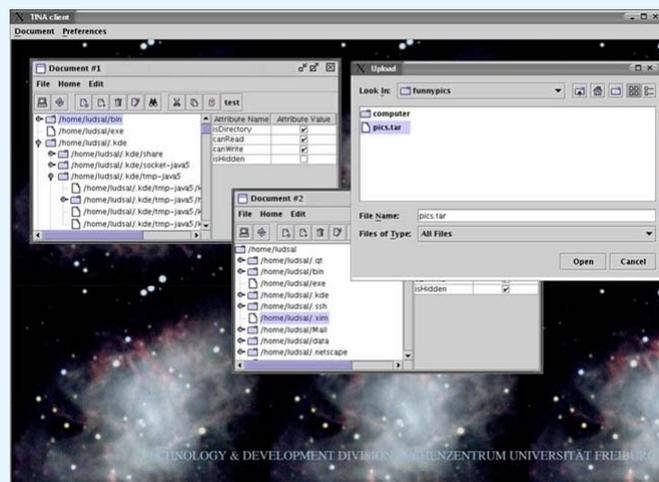
- Technical issues: archive management
  - Who owns how much data?
    - Run accounting for a faculty and charge.....
    - Automatic expiry of archive data (after 10 years, say)
  - Who owns what?
    - Rights management
  - Machine transparency
    - Store data from Windows and restore to Unix.....
- Volume issues:
  - Try to move 500 TB from old media to new
    - Time is becoming a very critical issue

## Archival issues

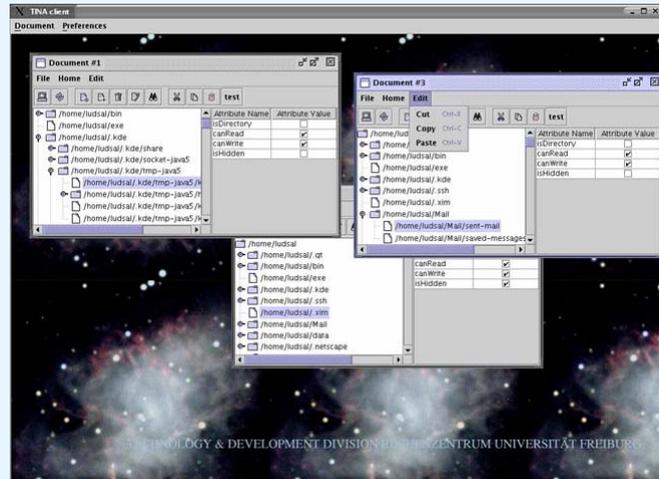
- Need an interface
  - to move data to archive space
  - to add additional information to the data
  - to control intrinsic rights
  - to ensure a long term data format
    - *.tif* is okay, *.doc* probably not
- Ongoing research project in Freiburg
  - Yet another digital library?  
Yes – but with a different aim



## Archival project



# Archival project



# Archival issues



- What about „living data“?
  - i.e. data that requires execution in a certain environment
  - like computer games
    - will historians of the future be able to understand what we did?
  - like multimedia presentations
    - PhD-thesis with animation
  - like computer algebra systems
    - to verify whether their bugs were used in „proofs“

## Living Data

- Ongoing research project
  - Provide an environment (description) so that programs can be executed on modern hardware
  - Ensure that the environment can be moved into the future – and test it on the past
    - First success: physics data became reusable
  - Execution can be automated – and thus made to fit the needs of libraries
- Whom should we give the data? TSM?

## Living Data ☺



Sinclair ZX

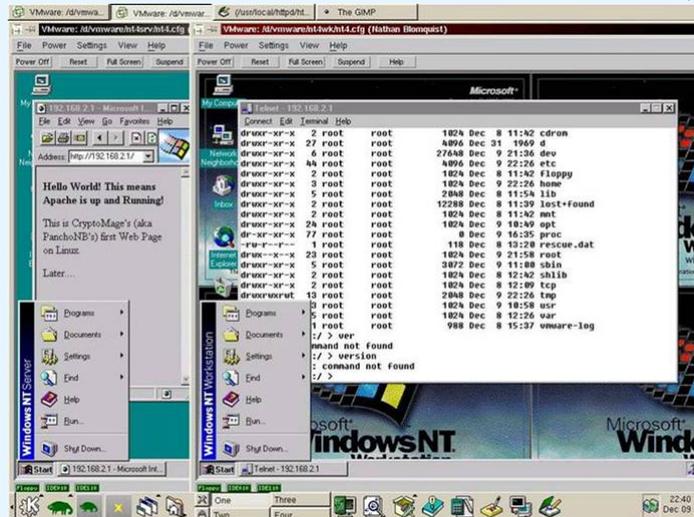


Commodore C64



Atari

# Living Data – Linux running Windows



## Conclusion:

TSM still has plenty of options for further development

