

Migration: Surfing on the Wave of Technological Evolution An ENSTORE Story



C.-H. Huang, W. Baisley, J. Bakken, D. Berg,
E. Berman, T. Jones, D. Litvinsev, A. Moibenko,
G. Oleynik, D. Petravick, G. Szmuska, M. Zalokar

Fermilab, Batavia, IL 60510, USA



Introduction

- ENSTORE

- petabyte scale permanent storage system
- tape based system
- developed by and installed at Fermilab
- highly available
- in service for 8 years
- 36,000+ volumes, 15,000,000+ files, 3.5 petabytes and counting



Nature of a Tape Based MSS

- Cost structure
 - total capacity = number of media X density
 - online capacity is limited by library
 - library is limited by physical facility
 - everything is limited by \$\$\$
 - *increase density, increase total capacity*



Nature of a Tape Based MSS

- Performance
 - maximal system performance is limited by the sum of individual drives' performance
 - increased drive performance increase overall system performance
- *Go for high density media and high performance drives*



Challenge for Long Running Production Systems

- Technologies evolve
 - newer media with higher density
 - better price/capacity
 - newer drives with better performance
 - better price/performance
- *Sticking with older technologies costs more!*



Taking Advantages of Newer Technologies

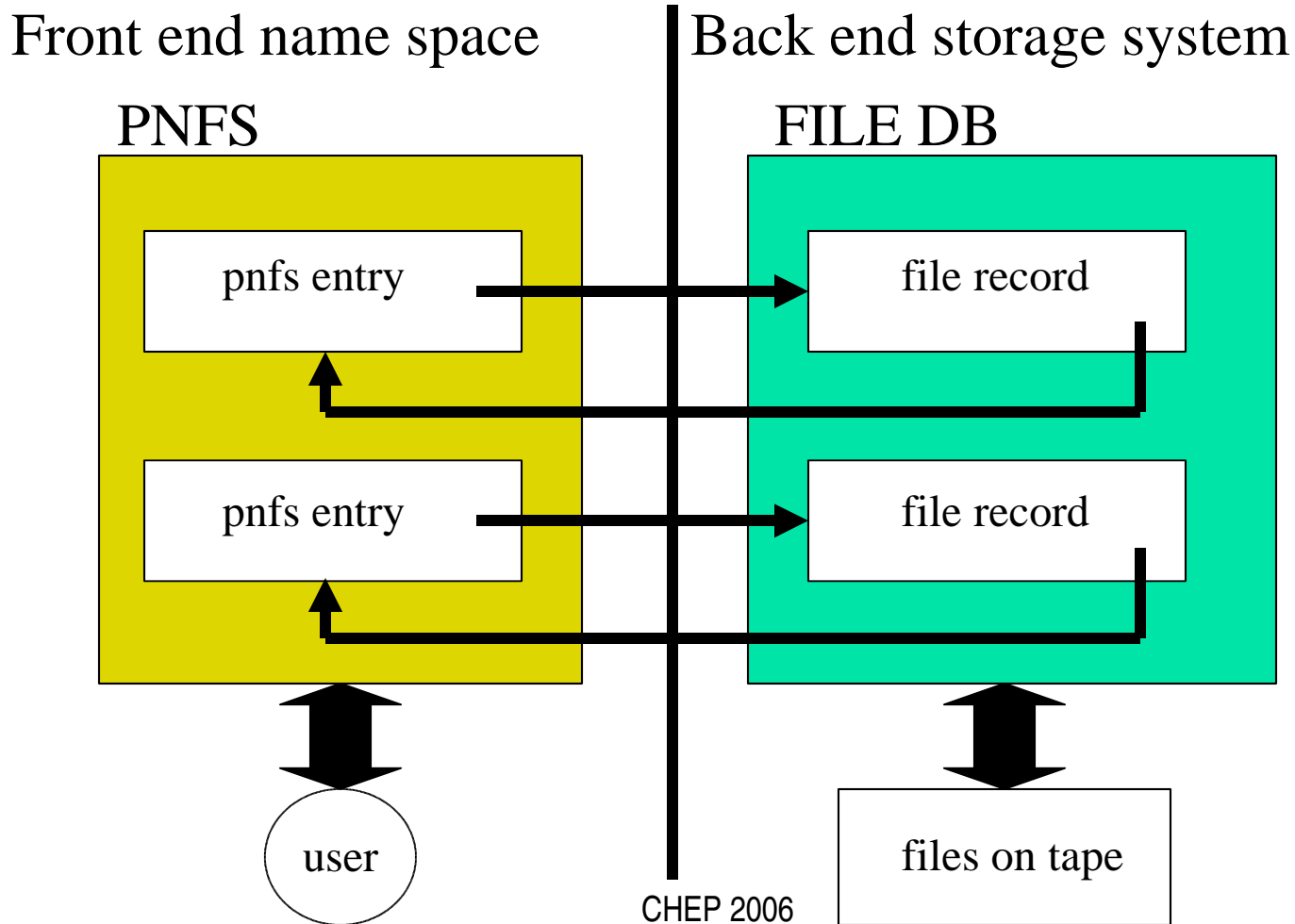
- Migrate data onto newer media
- Drive change often requires media change
- Example of 9940A to 9940B migration
 - same physical cartridge different formats
 - 3 times capacity increase (60GB to 200 GB)
 - 3 times performance increase (10MB/s to 30MB/s)
 - Migrating data from 9940A tapes to 9940B tapes and recycling/reformatting 9940A tapes to 9940B format increase the capacity and performance by a factor of three under the same physical and fiscal constraints



Conceptual View of ENSTORE

- Front end name space
 - Provides a file system like interface to users
 - Pnfs is the current implementation
- Back end storage system
 - Self sufficient with internal database for metadata
- Loosely coupled front end and back end
- Same metadata is stored in both places
- Deleted files are never removed unless the media is recycled

Conceptual View of ENSTORE





Design Considerations

- File based migration
- Files are always available
- Transparent to users
- No reserved resources for migration
- Minimal impact to system performance
- Complete transaction history log
- Minimal administrator attention
- Concurrent migration streams
- Reversible

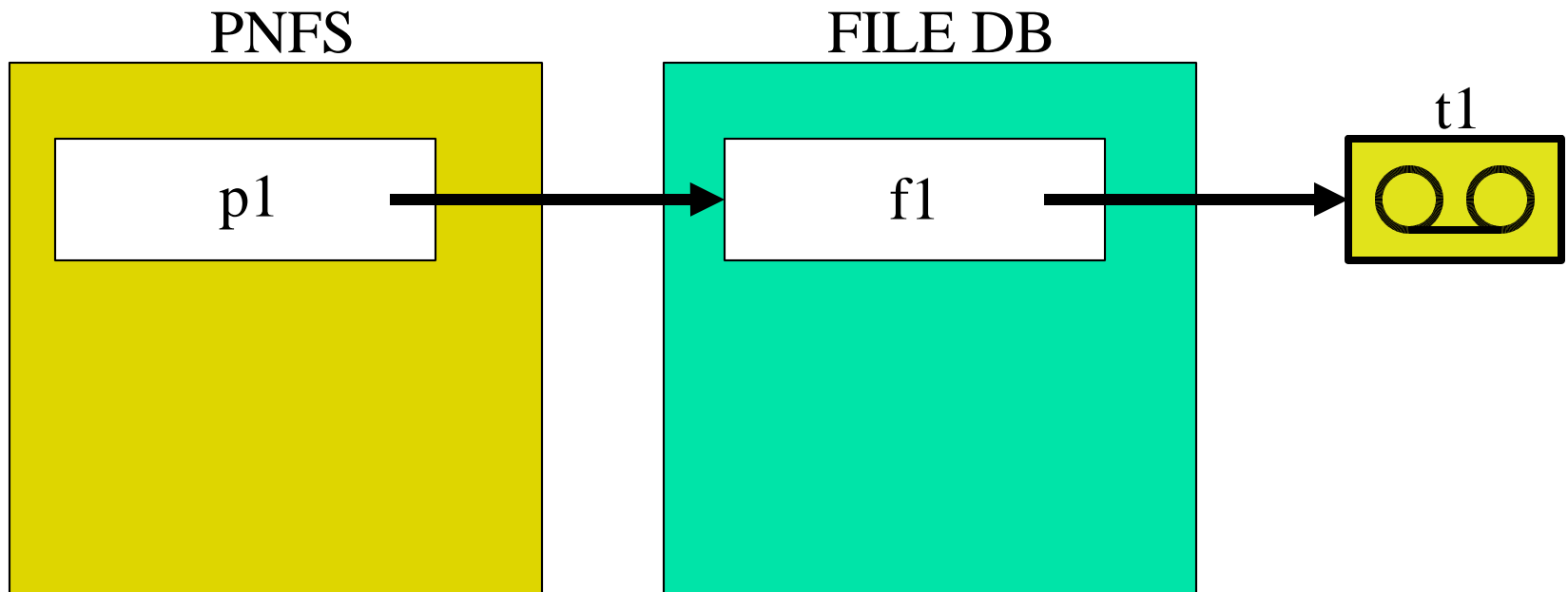


Implementation

- File migration in three steps
 - Copying file
 - Swapping metadata
 - Read back verification
- Batch mode (by tape)
 - As above
 - Deferred verification
 - More complicated error handling
- Optionally migrate deleted files

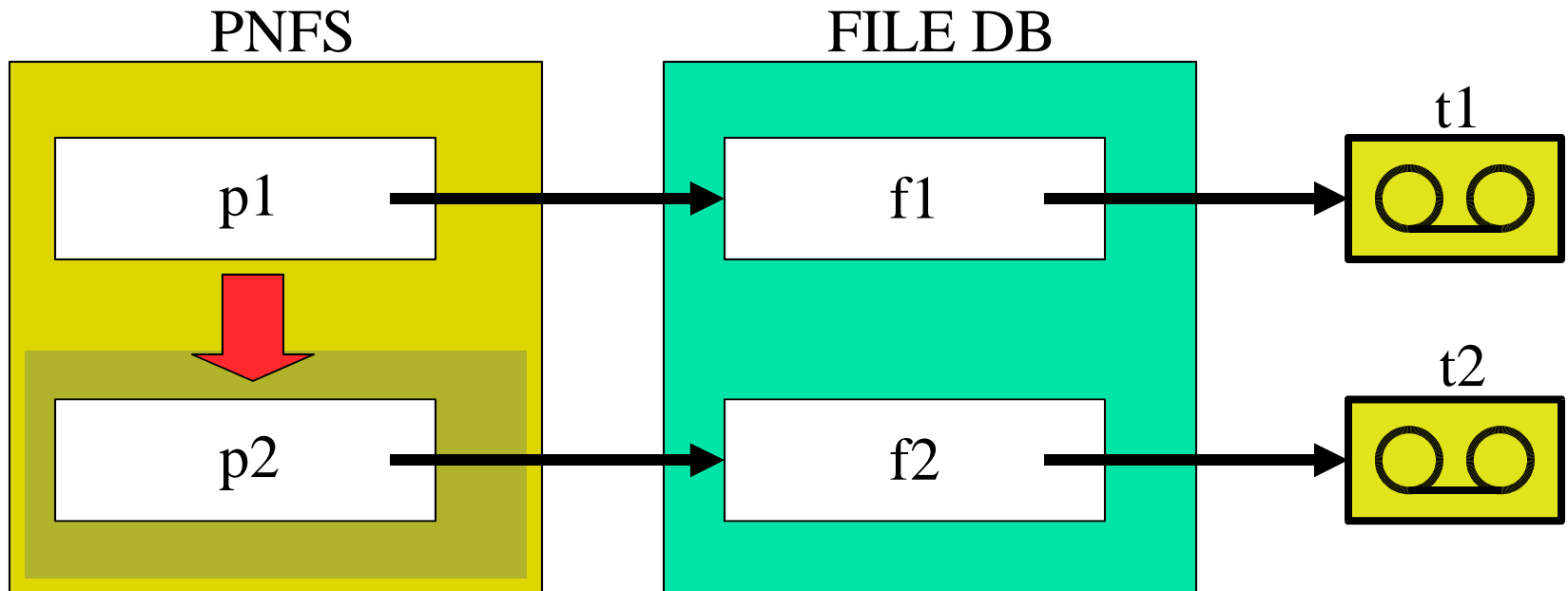
Before migration

- Users access file f1 through pnfs entry p1



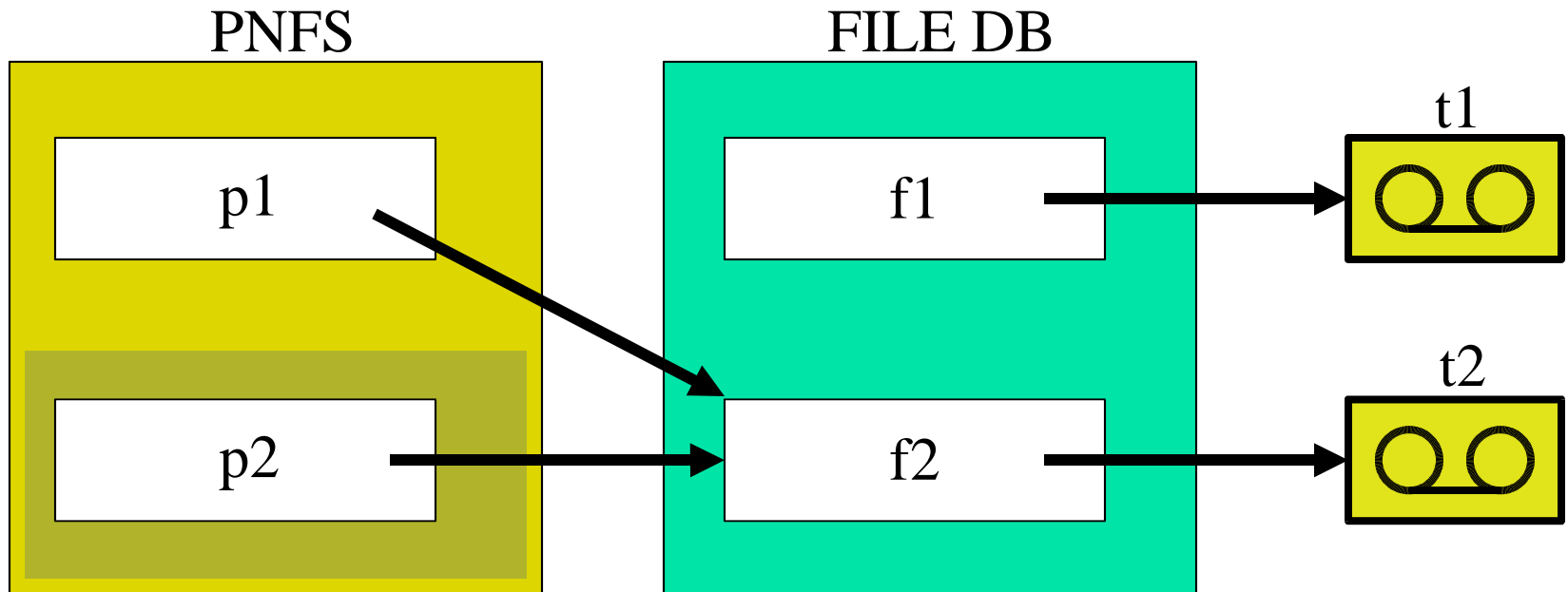
Step 1: Copying file

- file p1/f1/t1 is copied through disk to file p2/f2/t2
- f1 and f2 are distinct files of same content



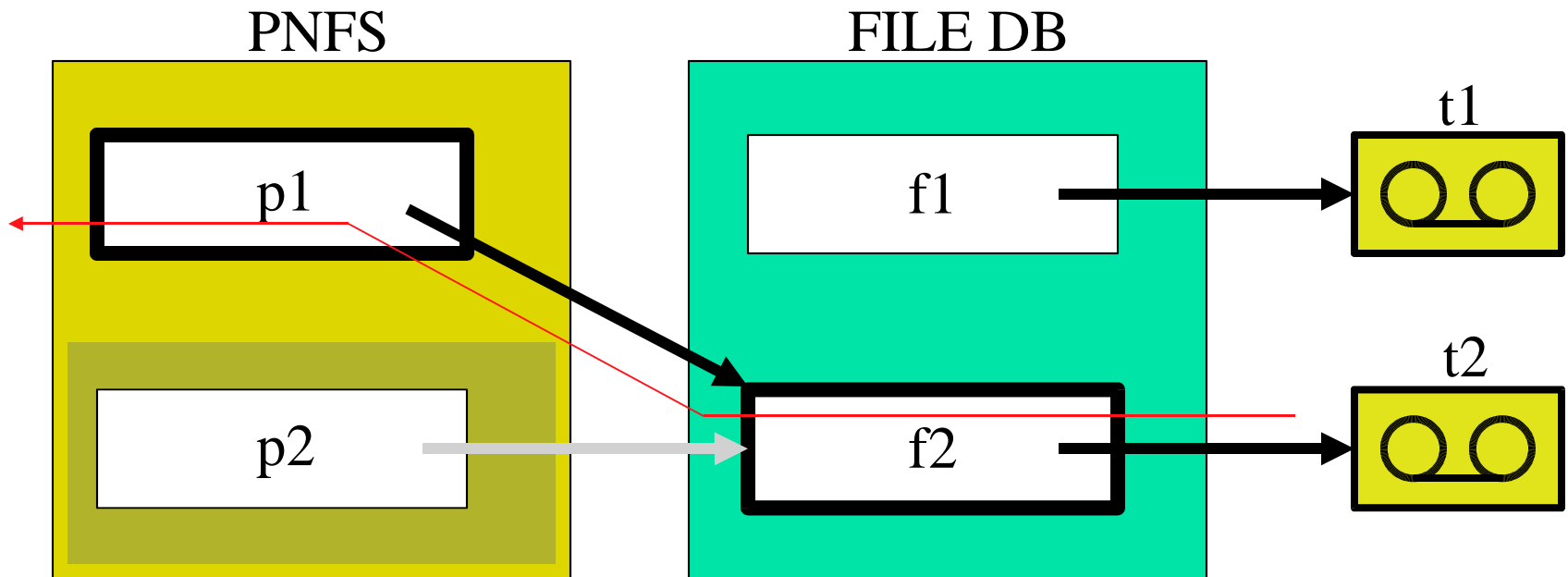
Step 2: Swapping Metadata

- a pair of one way copies
- f2 is immediately accessible



Step 3: Verification

- p1 reads f2
- f1 is marked deleted and, if there is a need, can be restored





Other Applications

- Media cloning
 - Cloning tapes without reserved resources
- Media compaction
 - Reclaiming space occupied by deleted files
- Media consolidation
 - Combining partially filled media to fewer ones



Experience and Concluding Remarks

- Migration in production for more than 2 years
 - Success story
- 2004: 4557 9940A tapes to 9940B migration
 - Triple the library capacity
- 2005: 1240 eagle tapes to 9940B migration
 - Reclaim 1000 needed slots in the library
- Migration becomes a routine task in ENSTORE
 - Keeps surfing on the wave of technological evolution