

High Performance Storage System Road Map



11/12/2007

Harry Hulen - hulen@us.ibm.com

Jim Gerry - jgerry@us.ibm.com

Background:

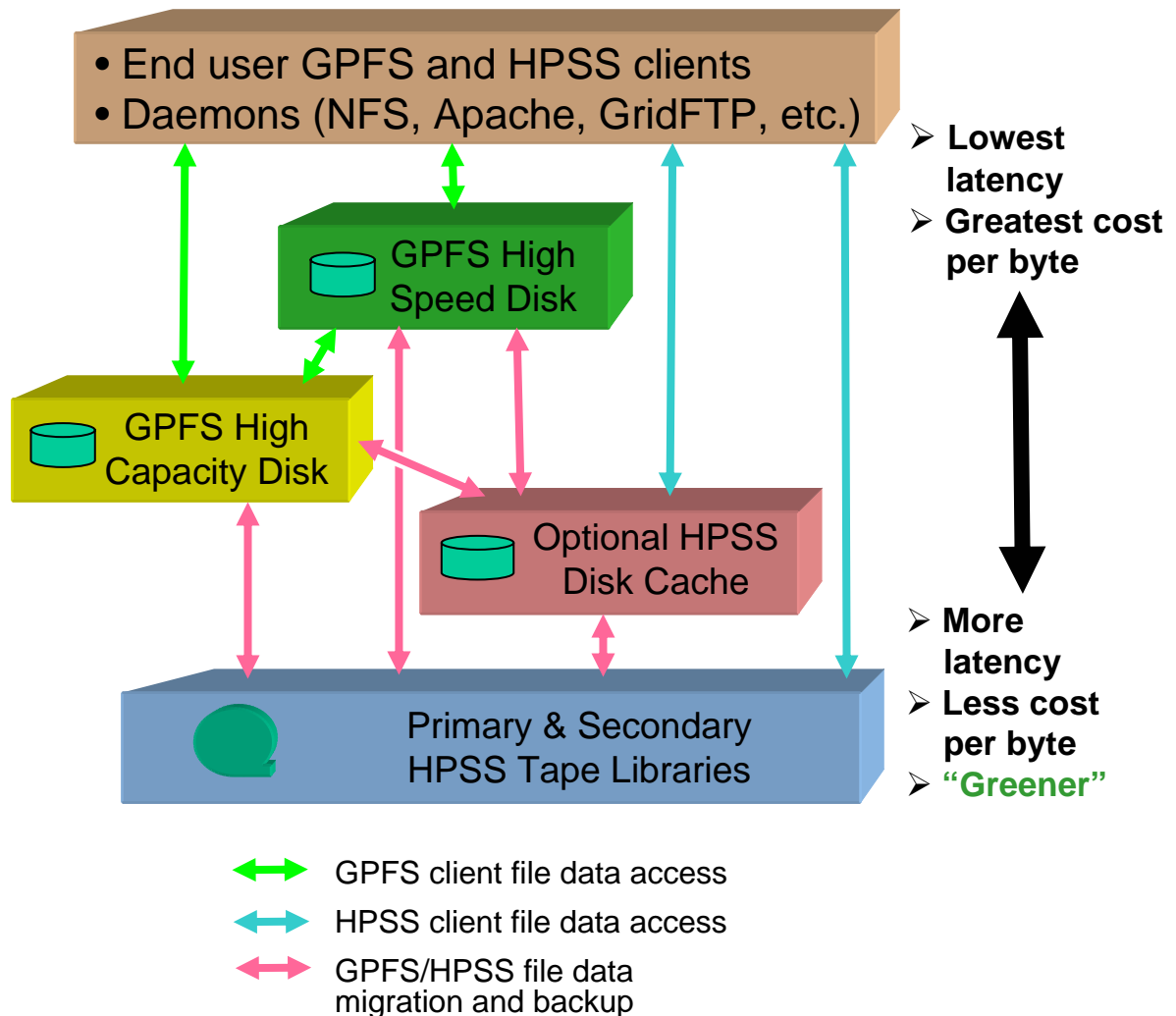
GPFS/HPSS Interface (GHI)

- A collaborative project of IBM's HPSS project, IBM Almaden Research Center, and Lawrence Berkeley National Lab
- Connect the GPFS and HPSS architectures together under the GPFS Information Lifecycle Management (ILM) policy framework and Data Management API (DMAPI)
- Provide a hierarchical GPFS file system having virtually unlimited storage capability
- Use the combined capabilities of GPFS and HPSS to provide disaster recovery protection for the GPFS file systems

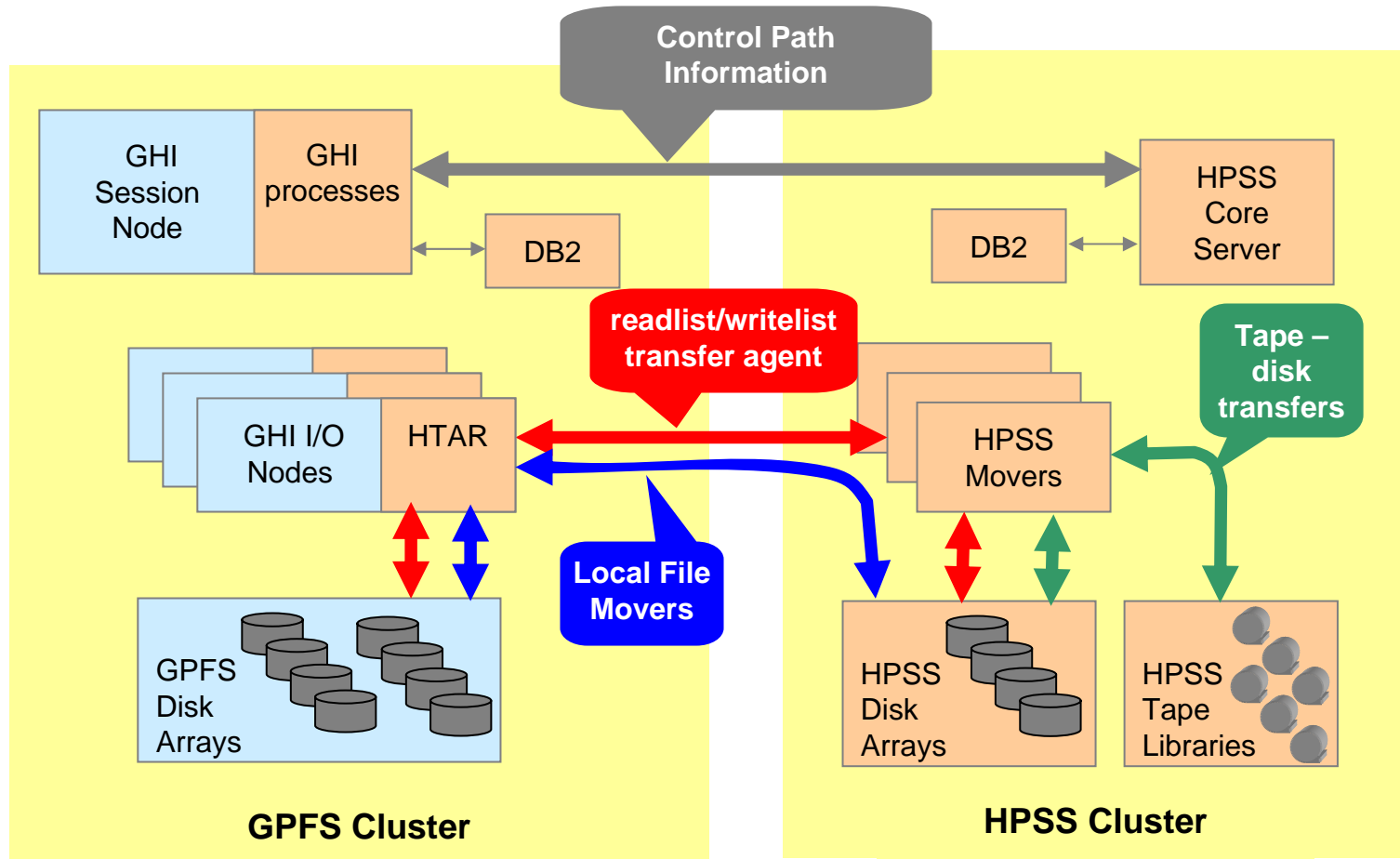
GPFS and HPSS

File Data Life-cycle Management

- Manages data and media over their life cycle.
- Migration of data and media is automatic and transparent to the client.
- Migration to HPSS is governed by GPFS rule-based ILM policy.
- Migration within HPSS is based on proven HPSS hierarchical storage management policy.
- HPSS provides both efficient space management and backup of GPFS.



GPFS/HPSS Data Movement with GHI



Flexible, Rule-driven Control for File Life-cycle Management

- Initial placement

```
RULE 'SlowDBase' SET STGPOOL 'sata' FOR FILESET('dbase') WHERE NAME LIKE '%.data'  
RULE 'SlowScratch' SET STGPOOL 'sata' FOR FILESET('scratch') WHERE NAME LIKE '%.mpg'  
RULE 'default' SET STGPOOL 'system'
```

Rule name

Storage pool name
(corresponds to
HPSS storage
class)

Fileset name
(corresponds to
subdirectory)

Qualifiers

- Movement by age

```
RULE 'MigData' MIGRATE FROM POOL 'system' THRESHOLD(80,78)  
WEIGHT(TIME_SINCE_LAST_ACCESS ) TO POOL 'sata' FOR FILESET('data')  
RULE 'HsmData' MIGRATE FROM POOL 'sata' THRESHOLD(95,80)  
WEIGHT(TIME_SINCE_LAST_ACCESS ) TO POOL 'hsm' FOR FILESET('data')  
RULE 'Mig2System' MIGRATE FROM POOL 'sata' WEIGHT(ACCESS_TIME) TO POOL 'system' LIMIT(85)  
FOR FILESET('user','root') WHERE DAYS_SINCE_LAST_ACCESS_IS_LESS_THAN( 2 )
```

Rule
to
move
files
to
HPSS

- Lock in place

```
RULE 'ExcDBase' EXCLUDE FOR FILESET('dbase')
```

- Life expiration

```
RULE 'DelScratch' DELETE FROM POOL 'sata' FOR FILESET('scratch') WHERE  
DAYS_SINCE_LAST_ACCESS_IS_MORE_THAN( 90 )
```

GPFS/HPSS Interface

History and Near-term Releases

- SC05 Conference – GHI concept demonstrated
- SC06 Conference – Production-ready GHI DMAPI version demonstrated
- SC07 Conference – New higher performance GHI incorporating small file aggregation and GPFS backup and restore demonstrated
- December 2007 – GHI SC07 capabilities generally available
- March 2008 – GHI SC07 capabilities production ready

Planned HPSS Releases

- HPSS 6.2.2
 - New GPFS/HPSS Interface beta release
 - Basis for Billion File Demo with GPFS
 - December 2007
- HPSS 6.2.2.1
 - New GPFS/HPSS Interface production release
 - March 2008
- HPSS 7.1
 - Major HPSS release
 - Features new HPSS native small file architecture
 - September 2008

History: HPSS 6.2

September 2006

- HPSS 6.2 is the **current major release**
- September 2006
- Current minor release is 6.2.1 (service pack 1)
- HPSS 6.2 was the final step in a three-year conversion
 - from the original Distributed Computing Environment (DCE) infrastructure
 - to a legacy-free infrastructure centered on DB2.
- First fully Linux-capable release
- Introduced POSIX-compliant VFS interface
- Enabled GridFTP, Apache (Web), Samba, and other 3rd-party interfaces

Goals of HPSS 6.2.2

December 2007

- Launch new GPFS/HPSS Interface (GHI)
- New hardware, OS, and DB2 support
- Provide fixes for 6.2 as needed

Details of HPSS 6.2.2 Platform Support

- Core computers:
 - AIX 5.3 on Power
 - RHEL 4 on Power and x86-64
 - 32-bit application on 64-bit kernel only
- Mover computers:
 - AIX 5.3 on Power
 - RHEL 4 on Power and x86-64
 - IRIX 6.5
 - 32-bit or 64-bit kernel
- Client computers:
 - All interfaces on RHEL 4
 - All interfaces except VFS on AIX 5.3, Solaris 9, IRIX 6.5
 - Client API and PFTP on RHEL 5
 - PFTP only on Solaris 10 on x86-64
 - 32-bit or 64-bit kernel

Details of HPSS 6.2.2 GPFS/HPSS Interface

- Ingest GPFS data into HPSS
- Stage/Recall data from HPSS to GPFS
 - DMAPI and ILM policies
- Provide a threshold policy to trigger an event to free data blocks in GPFS when the threshold is reached
- Provide file aggregation via HTAR to optimize small file performance
- Backup and restore the GPFS cluster & file system information
- Support GHI on AIX 5.3 and RHEL 4

Details of HPSS 6.2.2

Native HPSS Function

- New DB2 support:
 - DB2 V9.1 with 64-bit instance on core machines
 - DB2 data compression
- New tape drive support:
 - IBM LTO4 drives on AIX and RHEL 4
 - HP LTO4 drives on RHEL4

Goals of HPSS 6.2.2.1

March 2008

- Deliver additional GHI enhancements
 - Enhance performance and scalability
 - Support automatic staging of GPFS resident files after restore
 - Optimize staging data from tape
 - Enhance error handlings
 - Support symlinks and hardlinks
- Provide HPSS and GHI fixes for 6.2 as needed

Goals of HPSS 7.1

September 2008

- Introduce new native architecture for small files
- Increase scalability and performance
- Enhance ease of configuration and management
- Major OS upgrade
- Continue to enhance GPFS/HPSS Interface
- Provide a foundation for future enhancements

Details of HPSS 7.1 Platform Support

- Core computers:
 - AIX 6.1 on Power
 - RHEL 5 on Power and Intel x86-64
 - 32-bit application on 64-bit kernel only
- Mover computers:
 - AIX 6.1 on Power
 - RHEL 5 on Power and x86-64
 - 32-bit or 64-bit kernel
- Client computers:
 - All interfaces on RHEL 5
 - All interfaces except VFS on AIX 6.1
 - PFTP on Solaris 10
 - 32-bit or 64-bit kernel

Details of HPSS 7.1

Native Small File Enhancement

- Significantly improve metadata performance, especially when creating small files
 - Less and smarter locking on shared resources - (e.g. no locking on fileset metadata)
 - Less and smarter RPC usage - (e.g. create and open a file by a single call to Core Server).
 - Less and smarter DB2 usage - (e.g. batch updates of directory counts & access times, cache bitfile descriptor data across create/opens and maintain a cache of name objects)
- Aggregate small files on tape
 - Pack more files on a tape due to fewer tape marks
 - Enhance ability to maintain streaming speeds when writing small files
 - Applies to native HPSS files (GPFS/HPSS has its own)



Details of HPSS 7.1

Other Performance Enhancement

- Dynamically set the disk segment length for each file
 - Reduce number of segments to save metadata records
 - More efficient use of disk space when file size is not known ahead of time, as with POSIX write, copy
 - Start with smallest segment size
 - Double with each new segment till we hit max size
 - Final disk segment is always truncated for best fit
- Spread Mover workload by mover loading vs. today's round robin approach

Details of HPSS 7.1

Administrative Features

- Add or remove devices without recycling HPSS servers
- Most administrative functions now command line as well as GUI driven
- Change Classes of Service in large batches using multiple streams
- Ability to convert 6.2 metadata to 7.1 format concurrently with ongoing operations, significantly lessening downtime for upgrade

Details of HPSS 7.1 GPFS/HPSS Interface

- Performance enhancements
 - More intelligent load balancing across cluster
 - Optimize staging data
- Additional error handling
- Checksum capability to assure file content integrity

Preview of HPSS in 2009 – 2011

from Dick Watson, LLNL, co-chair of HPSS Executive Committee

