

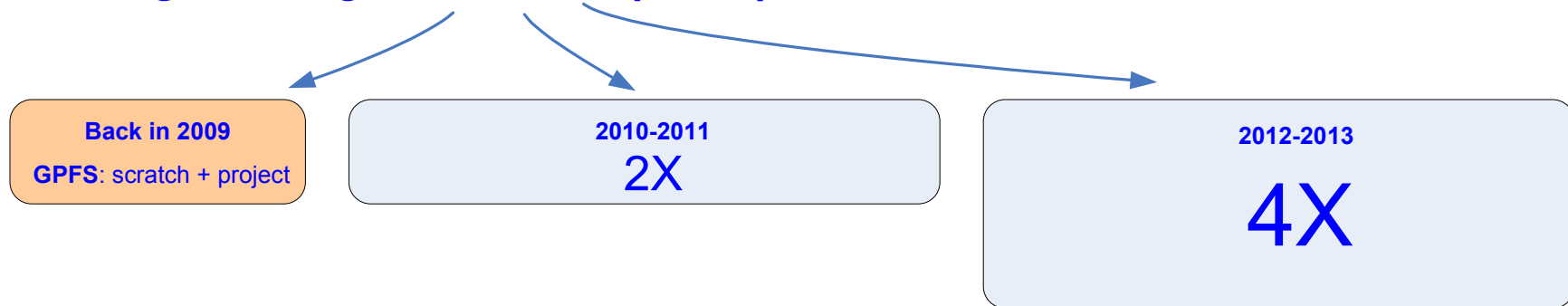
The World of HPSS



Storage Capacity Expansion Plan



Storage Budget: \$ \$ \$ (5PB)



Rational:

- * the longer we wait, the more resources we can buy with the same budget

Manage usage and expectation:

- * allocations
- * introduction of quotas & [HSM^{IBM}+TSM] for limited data offload
- * regular purging

What have we learned in 1 ½ years of operation?

- * GPFS problems and limitations at a 4000 nodes scale
- * User data distribution patterns not GPFS or HPC friendly

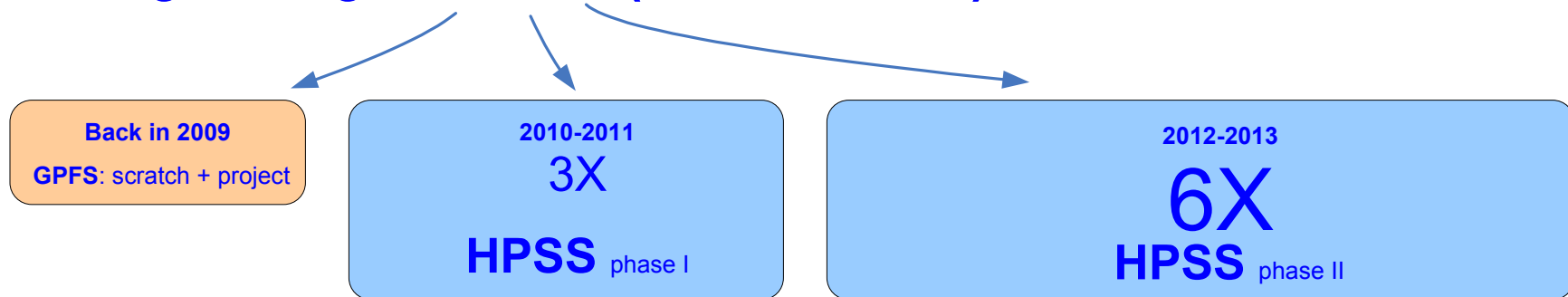
Conclusion:

- * just adding spinning disks to active filesystems is not a good solution
- * more users, more data, more files => more problems

Storage Capacity Expansion Plan



Storage Budget: \$\$\$ (5PB or more)



Solution:

- * near online storage with HPSS (tape backed hierarchical storage system)

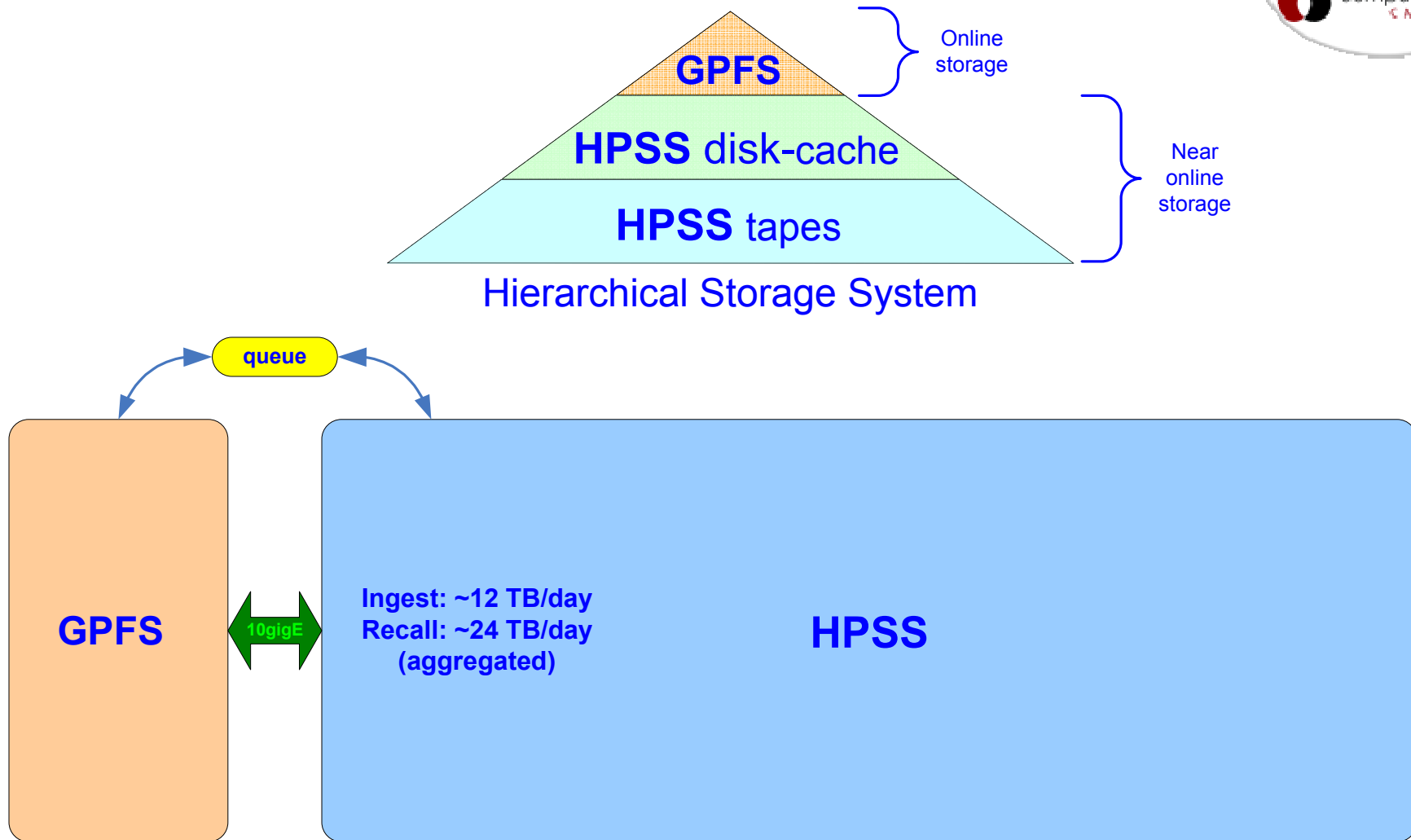
Manage usage and expectation:

- * allocations: GPFS + HPSS
- * quotas & massive offload data to HPSS
- * regular purging
- * less utilization of small files
- * more utilization of tarballs in the regular workflow by users

About HPSS (High Performance Storage System):

- * 10+ years history, used by 50+ facilities in the "Top 500" HPC list
- * very reliable, data redundancy and data insurance built-in.
- * highly scalable, reasonable performance at SciNet
- * HSI/HTAR clients also very reliable and used on several HPSS sites

Storage Capacity Expansion Plan



- * access and transfer management through the Queue System
- * end-user interaction via HSI/HTAR calls in the job scripts

Scripted File Transfers



File transfers in and out of the HPSS should be scripted into jobs and submitted to the archive queue. Scripts should use the HSI and/or HTAR commands as in the example below:

```
#!/bin/env bash
#PBS -q archive
#PBS -N hsi_put_file_in_hpss
#PBS -j oe
#PBS -me

/usr/local/bin/hsi -v <<EOF
cput -p /scratch/$USER/workarea/finished-job1.tar.gz : finished-job1.tar.gz
EOF
status=$?
if [ ! $status == 0 ];then
    echo '!!! TRANSFER FAILED !!!'
fi
exit $status
```

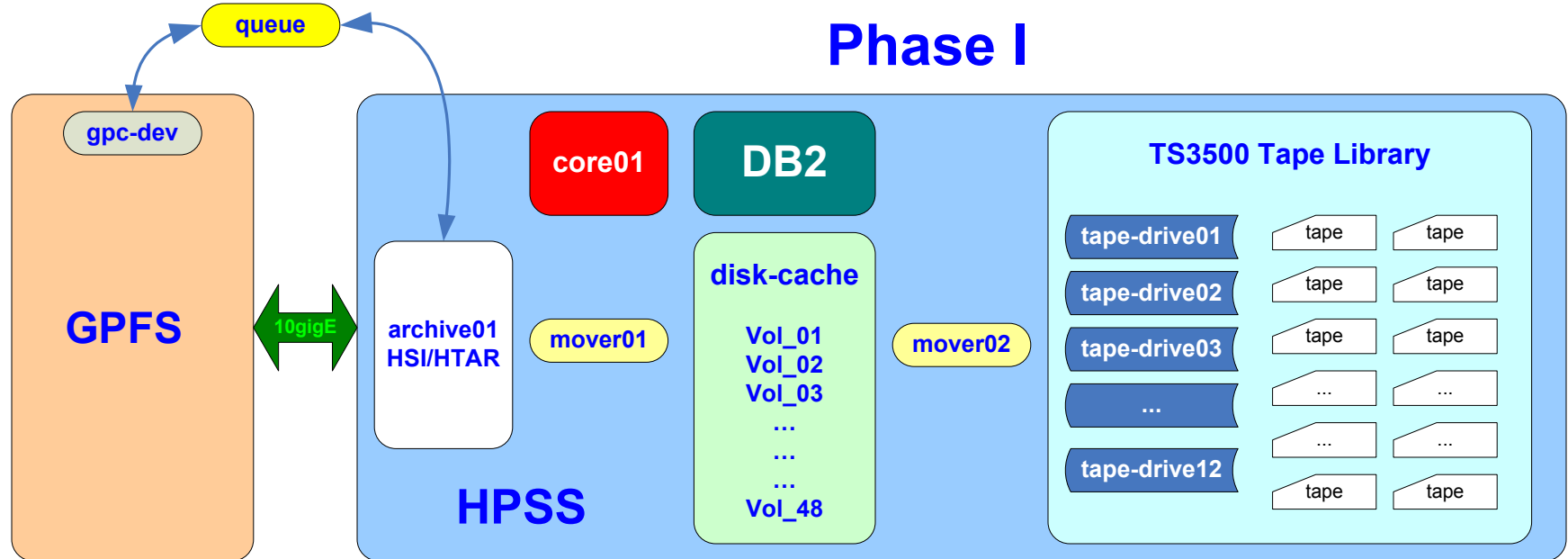
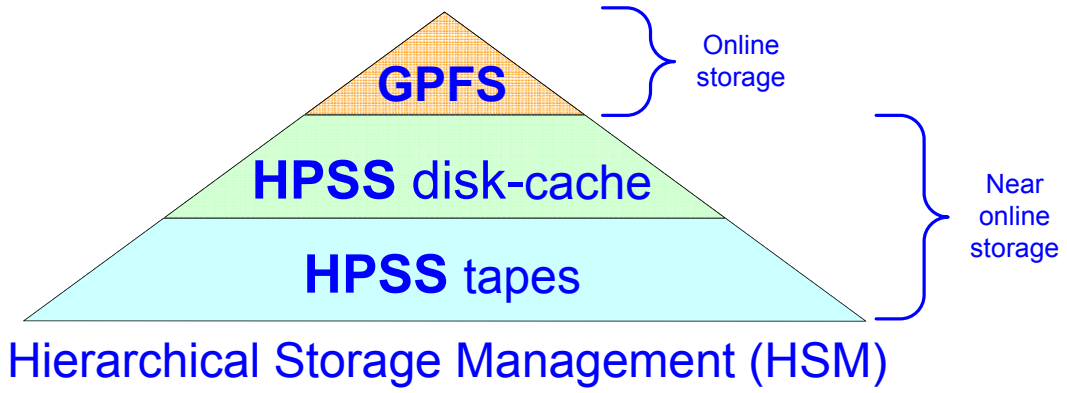
Note: Make sure to check the application's exit code and the returned log file for errors after all data transfers and any tarball creation process

The status of pending jobs can be monitored with showq specifying the archive queue:

```
showq -w class=archive
```

For more details please consult the the following wiki page:
<https://support.scinet.utoronto.ca/wiki/index.php/HPSS>

HPSS – main components



HPSS* =
 nodes + disks + network + HPSS + DB2 + HSI + HTAR + Library + tapes + services

HPSS – scaling potential

Phase II

