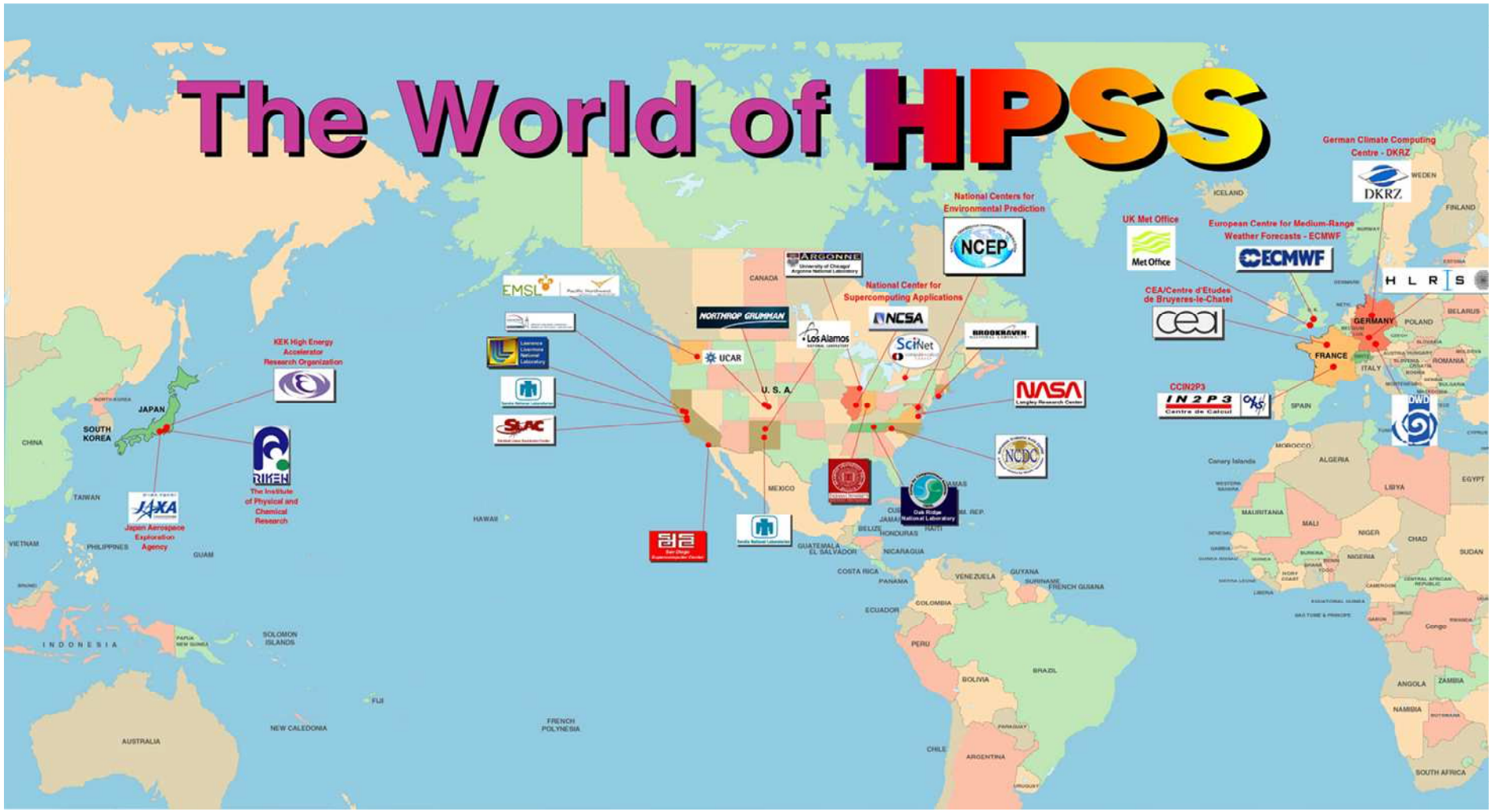
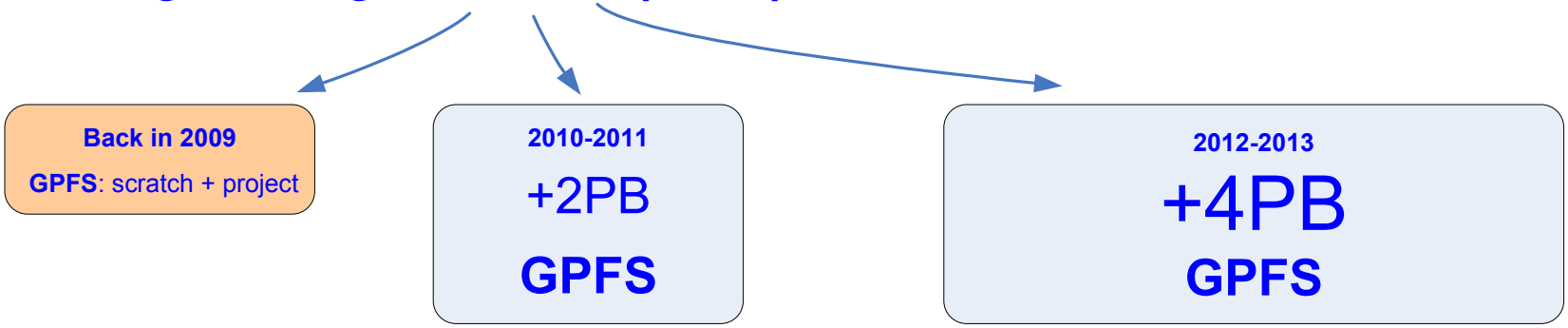

The World of HPSS
National Centers for Environmental Prediction
NCEP
German Climate Computing Centre - DKRZ
DKRZ
UK Met Office
Met Office
European Centre for Medium-Range Weather Forecasts - ECMWF
ECMWF
National Center for Supercomputing Applications
NCSA
CEA/Centre d'Etudes de Bruyeres-le-Chatel
ceal
KEK High Energy Accelerator Research Organization
RIKEN
The Institute of Physical and Chemical Research
JAXA
Japan Aerospace Exploration Agency
CCIN2P3
IN2P3
Centre de Calcul
NASA
Langley Research Center
NCDC
Oak Ridge National Laboratory
EMSL
ARGONNE
University of Chicago Argonne National Laboratory
NORTHROP GRUMMAN
Los Alamos
BROOKHAVEN
SciNet
UCAR
SLAC
U. S. A.
CANADA
MEXICO
JAPAN
SOUTH KOREA
CHINA
TAIWAN
VIETNAM
PHILIPPINES
GUAM
HAWAII
INDONESIA
SOLOMON ISLANDS
FIJI
NEW CALEDONIA
FRENCH POLYNESIA
AUSTRALIA
GUATEMALA
EL SALVADOR
HONDURAS
NICARAGUA
COSTA RICA
PANAMA
COLOMBIA
VENEZUELA
GUYANA
SURINAME
FRENCH GUIANA
ECUADOR
PERU
BOLIVIA
BRAZIL
PARAGUAY
CHILE
ARGENTINA
URUGUAY
ICELAND
NORWAY
SWEDEN
FINLAND
ESTONIA
BELARUS
POLAND
GERMANY
FRANCE
SPAIN
ITALY
ROMANIA
MOROCCO
Canary Islands
ALGERIA
LIBYA
EGYPT
MAURITANIA
MALI
NIGER
CHAD
SUDAN
NIGERIA
Congo
ANGOLA
ZAMBIA
NAMIBIA
SOUTH AFRICA

Storage Capacity Expansion Plan (initial)

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



Rational:

* the longer we wait, the more we can buy with the same budget (hopefully)

Usage management:

* allocations
* introduction of quotas & HSM (limited data offload capability)
* regular purging

What have we learned in 1 ½ years of operation?

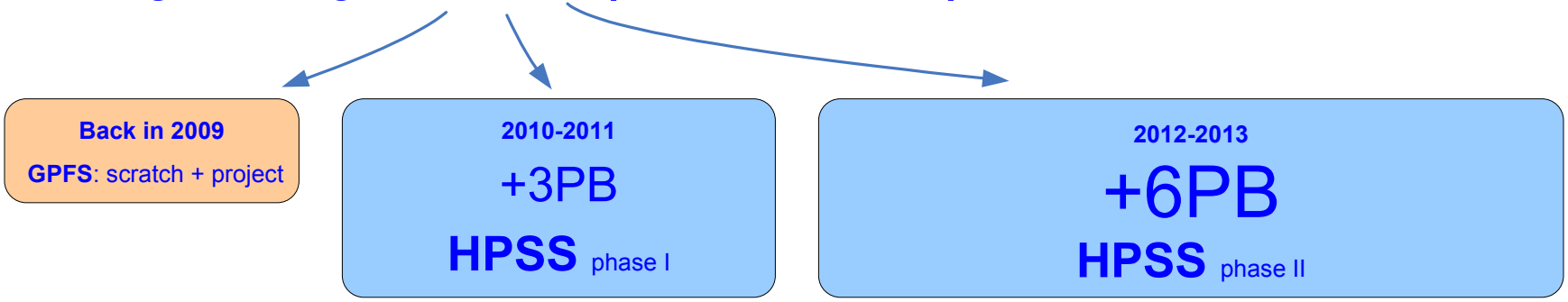
* GPFS problems and limitations at our scale (4000 nodes cluster)
* user data distribution patterns not GPFS or HPC friendly

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 (revised)

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



Solution:

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

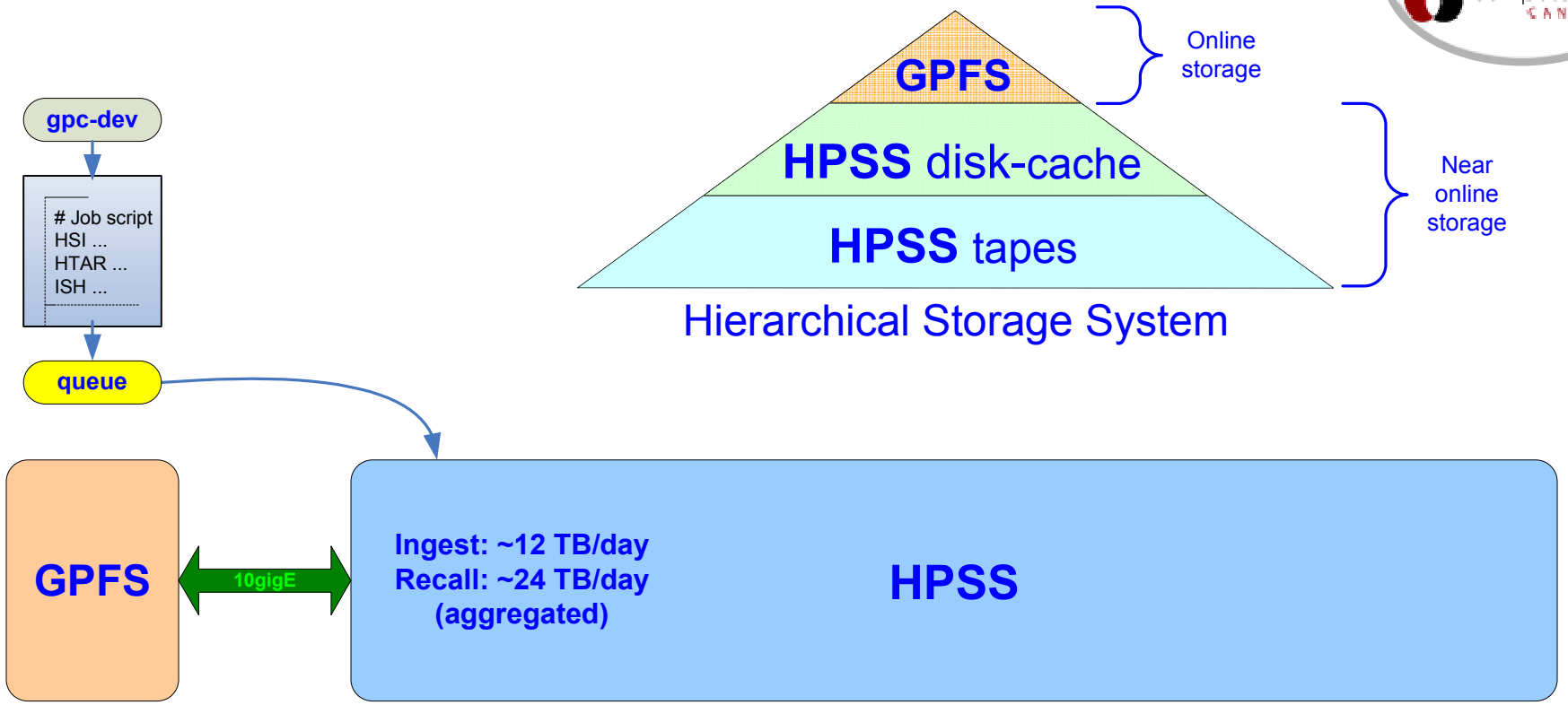
Usage management:

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

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 (and ISH) clients also very reliable and used on several HPSS sites.

How it works ...



* access and transfer management is done through the GPC queue system
* end-user interaction via HSI/HTAR/ISH calls in the job scripts

* HSI is a client with an ftp-like interface which can be used to archive and retrieve large files. It is also useful for browsing the contents of HPSS.
* HTAR is a utility that creates tar formatted archives directly into HPSS. It also creates a separate index file (.idx) that can be accessed quickly.
* ISH is a TUI utility to perform an inventory of contents in your tarballs.

Scripted File Transfers


```
#!/bin/bash
#PBS -q archive
#PBS -N htar_create_tarball_in_hpss
#PBS -j oe
#PBS -m e

echo "Creating a htar of finished-job1/ directory tree into HPSS"

trap "echo 'Job script not completed';exit 129" TERM INT
# Note that your initial directory in HPSS will be /archive/$(id -gn)/$(whoami)/

cd /scratch/$(whoami)/workarea/
htar -cpf /archive/$(id -gn)/$(whoami)/finished-job1.tar finished-job1/
status=$?

trap - TERM INT

if [ ! $status == 0 ]; then
   echo 'HTAR returned non-zero code.'
   /scinet/gpc/bin/exit2msg $status
   exit $status
else
   echo 'TRANSFER SUCCESSFUL'
fi
```

headers

htar
hsi
ish

trap

status

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

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

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

ISH used from the command line:

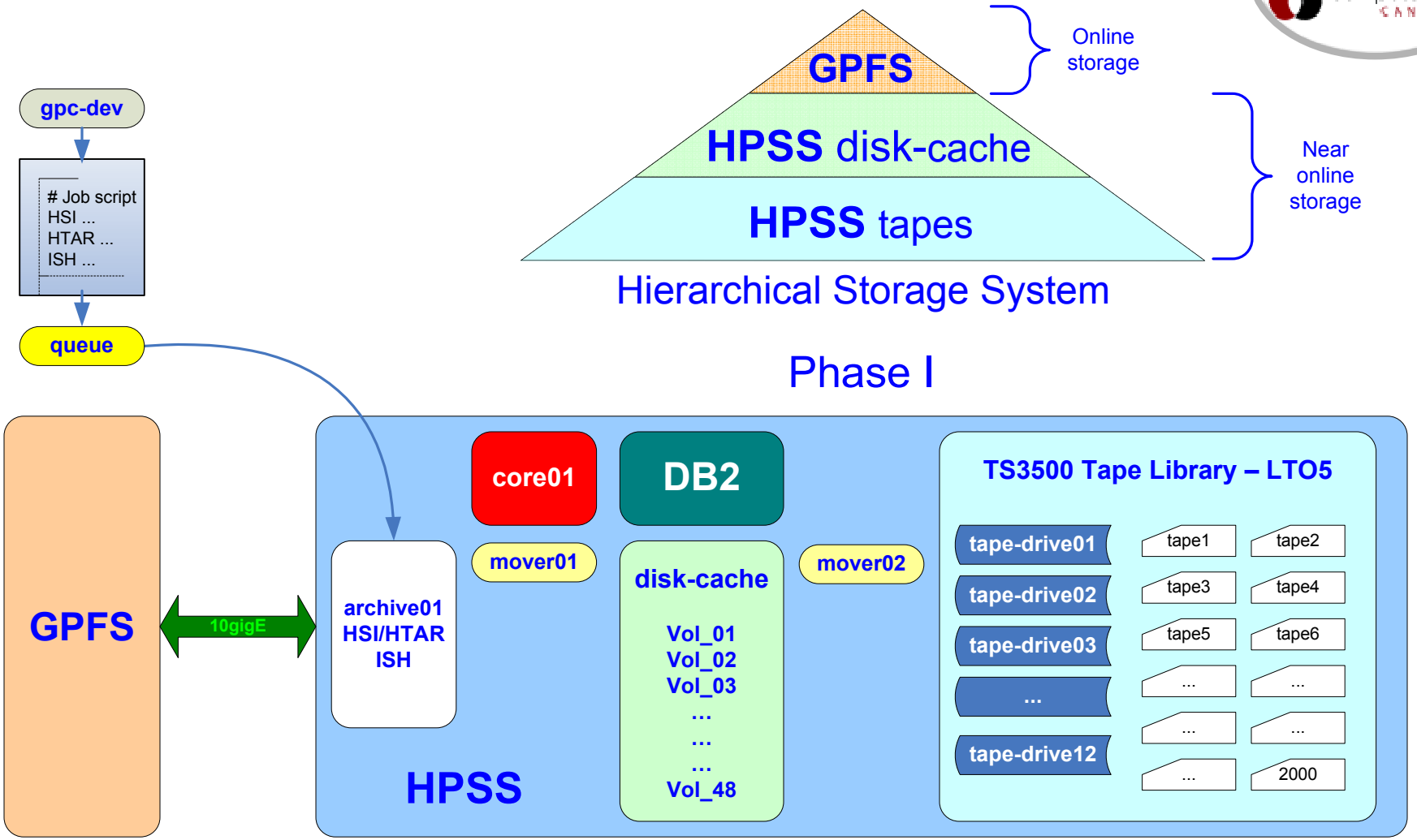
```
rzon@scinet02:~$ ls
data.tgz
rzon@scinet02:~$ /scinet/gpc/bin/ish
ish 0.98
Ramses van Zon - SciNet/Toronto/Canada/July 8, 2011
[ish]hpss.igz> index data.tgz
[ish]data.tgz.igz> ls -l
drwxr-xr-x rzon/scinet            0 2011-02-10 13:57:01 data/
-rw-r--r-- rzon/scinet        16714 2010-10-05 12:41:45 input.ini
-rwxr-xr-x rzon/scinet          293 2011-06-30 12:42:57 submit.pbs
[ish]data.tgz.igz> cd data
[ish]data.tgz.igz> ls
run1/   run2/
[ish]data.tgz.igz> find important*.dat
run1/important01.dat  run1/important02.dat  run1/important03.dat
run1/important04.dat  run1/important05.dat  run1/important06.dat
run2/important01.dat  run2/important02.dat  run2/important03.dat
[ish]exit
rzon@scinet02:~$
```

ISH used from a job script:

```
#!/bin/bash
# This script is named: data-list.sh
#PBS -q archive
#PBS -N hpss_index
#PBS -j oe
#PBS -m e
/scinet/gpc/bin/ish hindex
```

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

HPSS – main components



HPSS (broad use of the term) =
nodes + disks + network + FC + HPSS + DB2 + HSI + HTAR + ISH + Library + tapes + services

HPSS – scaling potential