



High Performance Storage System for long-term data stewardship

2023 DLR HSM Conference

Jim Gerry
Architect & Consultant
IBM Consulting



Today's agenda

- Why deploy storage tiers?
- Top three storage tier technologies.
- The value of tape.
- What is HPSS?
- HPSS is best of breed for long-term data stewardship.
- Converting to HPSS.

Why deploy different tiers of storage?

- Not all data are the same.
 - Small files vs. Large files.
 - Hot data vs. Warm data vs Cold data.
 - Individual files vs. Data sets.
 - Throw-away files vs. Save forever files.
 - Not important vs. Very important.

- In our experience...
 - Ninety-Ten Rule: 90% of files take up 10% of the space.
 - Files are only “active” or “hot” for a few days or months.
 - Folks like to group data.
 - Sets of data are often created at the same time.
 - Sets of data are often used at the same time.

.ico	.mov
.thumbnail	.tiff
.jpg	.tar
.tmp	.tiff
.prod	.raw

Many Small Few Big
 Hot → Cool



Not all storage is the same.

- Flash = over \$ 20x/TB
 - Highest performing.
 - Most expensive.

- Spinning disks = over \$ 9x/TB
 - Being squeezed by flash and tape.
 - Reached superparamagnetic limit.
 - Power-hungry.

- Tape = \$ x/TB
 - Highest latency.
 - Gold standard for data protection.
 - Lowest priced.



IBM FlashSystem
375 TB NVMe

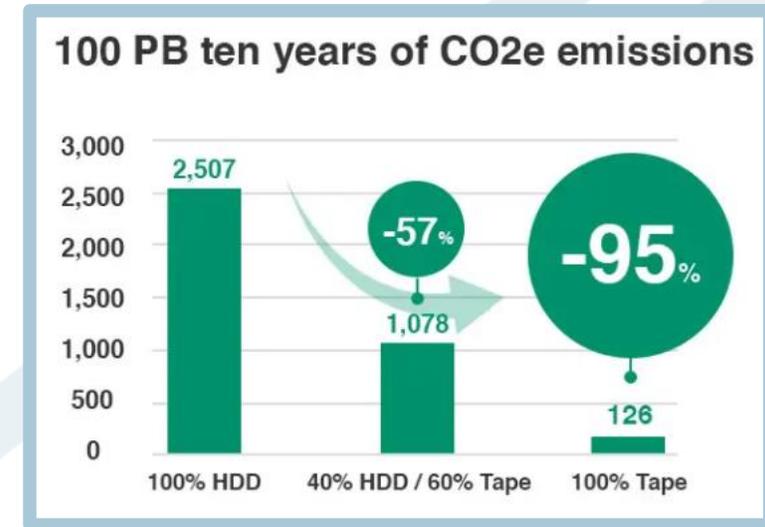
20 PB IBM Storwize

One TS4500 D-frame
740 JF tapes (37 PB)
16 Enterprise Drives



Value of tape for archive storage.

- Air-Gap protection – best defense against ransomware – Fuji Film
 - Hackers and malware can't get to your data.
- Tape is Near-Zero-Watt storage - Brad Johns Consulting
 - Moving 60% of data to tape cuts CO2 emissions by 57%.
 - Tape is GREEN – 95% lower carbon emissions vs. HDD.
- Cloud vendors understand the value of tiers and tape.
 - Cloud users leverage archive storage to SAVE MONEY!
 - Organizations leverage HPSS to do the same!
- The more data you put on tape...
 - The more money you have for Flash storage.
 - The more power you have for HPC.
- HDDs are best to cut random access latency to archive storage.





“Archive” – what comes to mind?



Screen grab from Raiders of the Lost Ark



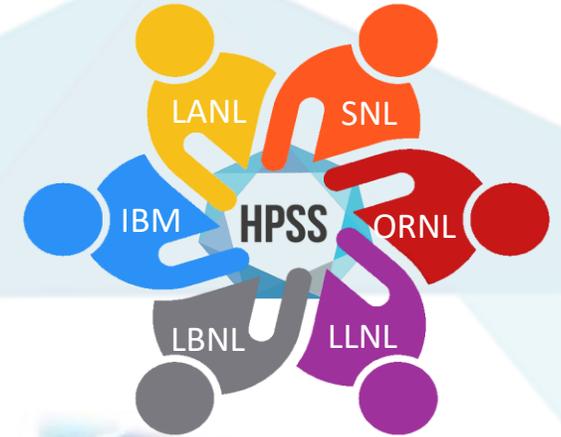
What is HPSS?

”IBM’s best kept secret”
”Best of breed for tape”

HPSS

What is HPSS?

- HPSS is storage software for those warm and cold tiers.
- HPSS is a 31-year long-running development collaboration between IBM and five US Department of Energy labs.
- HPSS is an IBM service offering from IBM Consulting.
 - Architect the solution.
 - Deliver and configure the software and provide training.
 - Demonstrate performance and functionality.
 - Deliver personalized support – a relationship.
 - **No capacity-based license for disk and tape storage.**
- HPSS is not an IBM storage product.

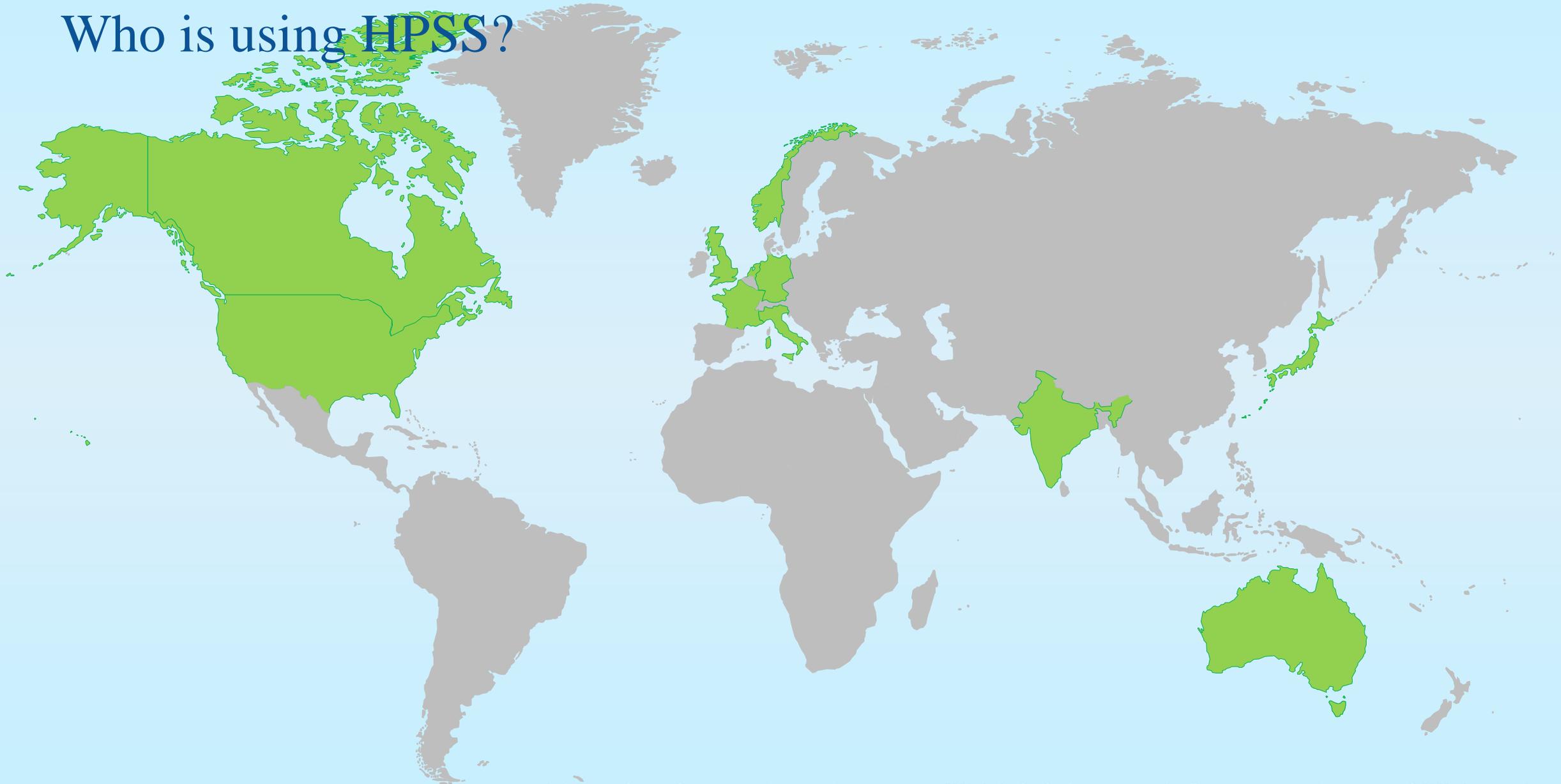


HPSS



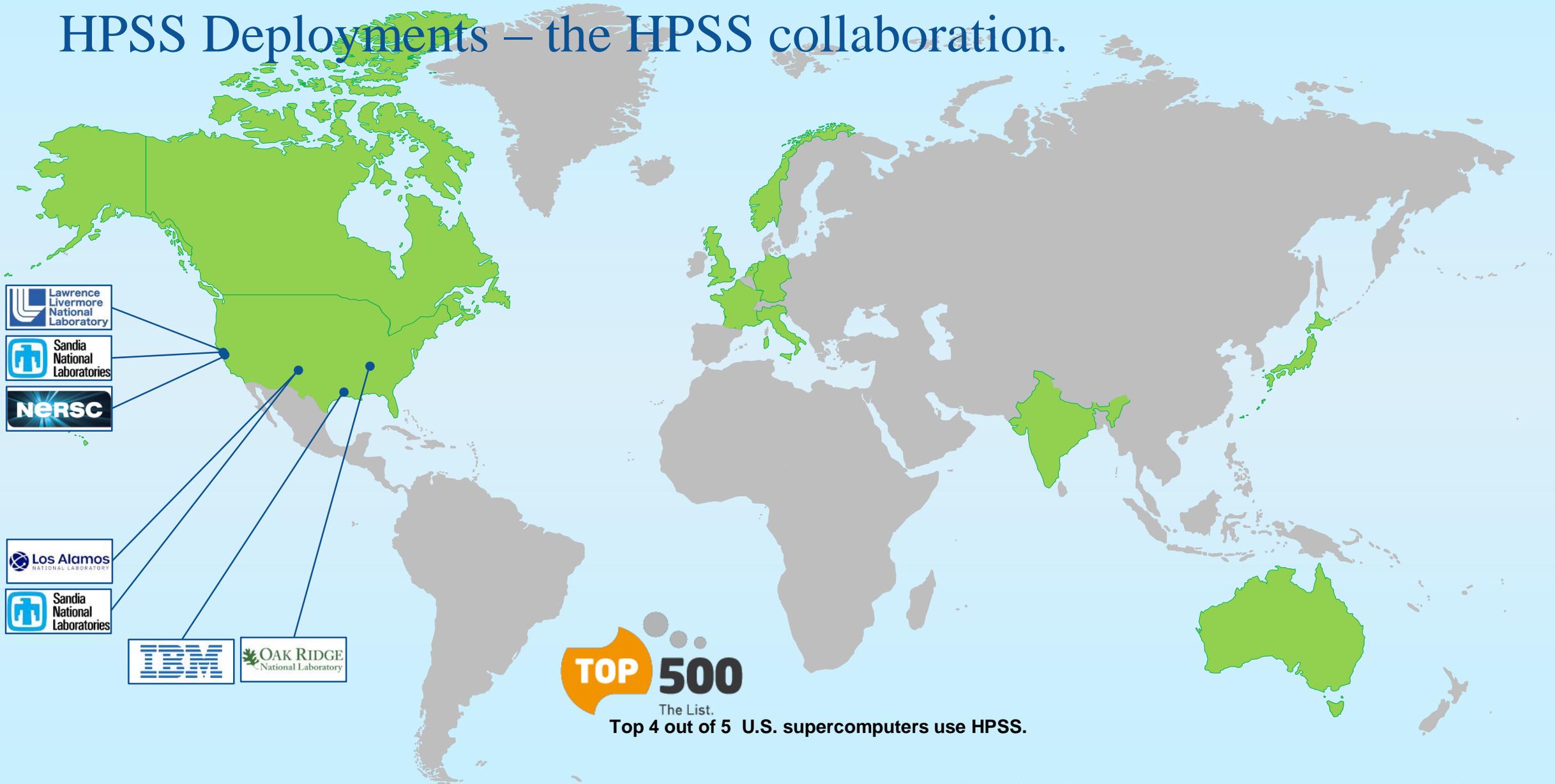


Who is using HPSS?

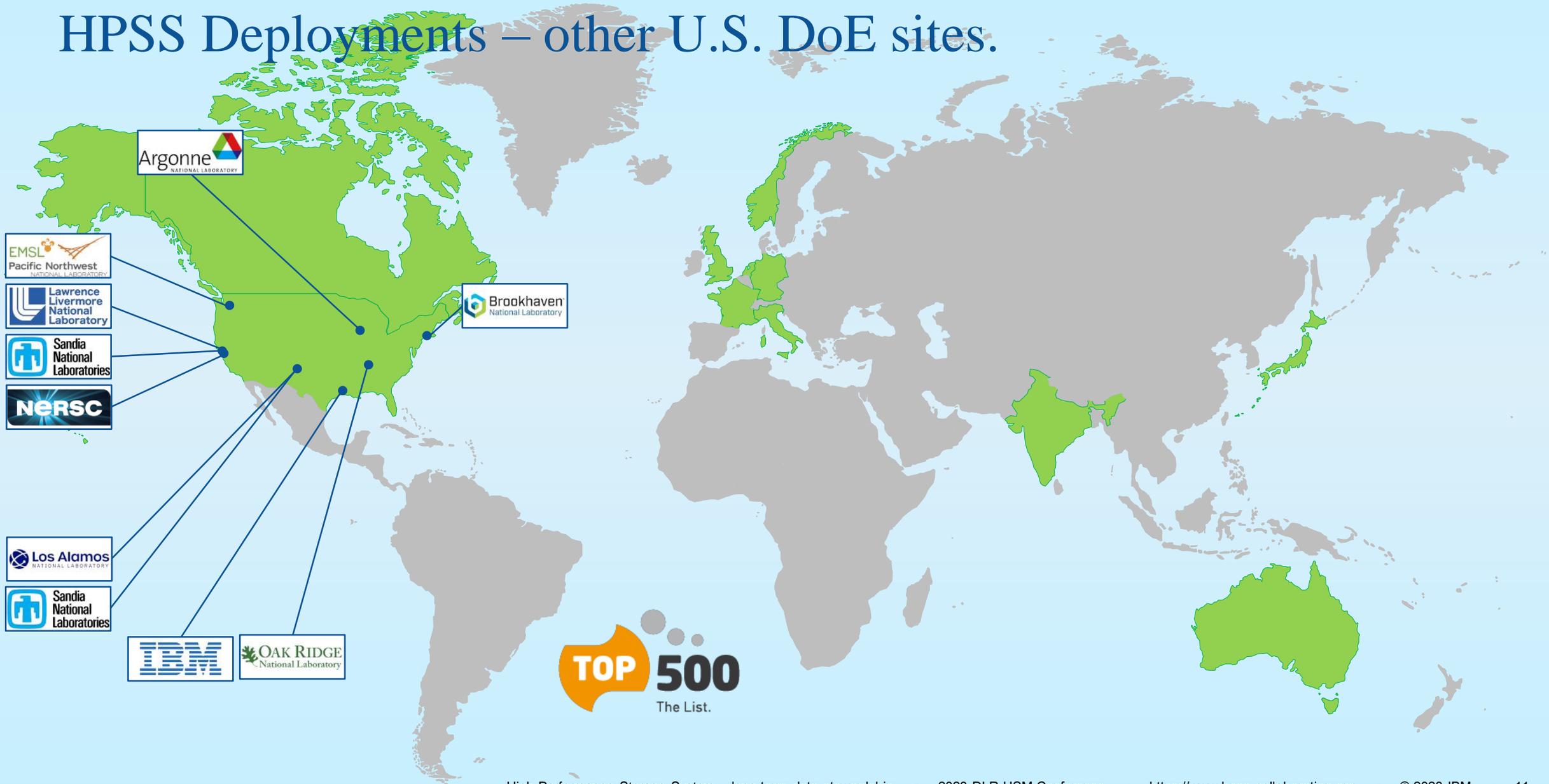




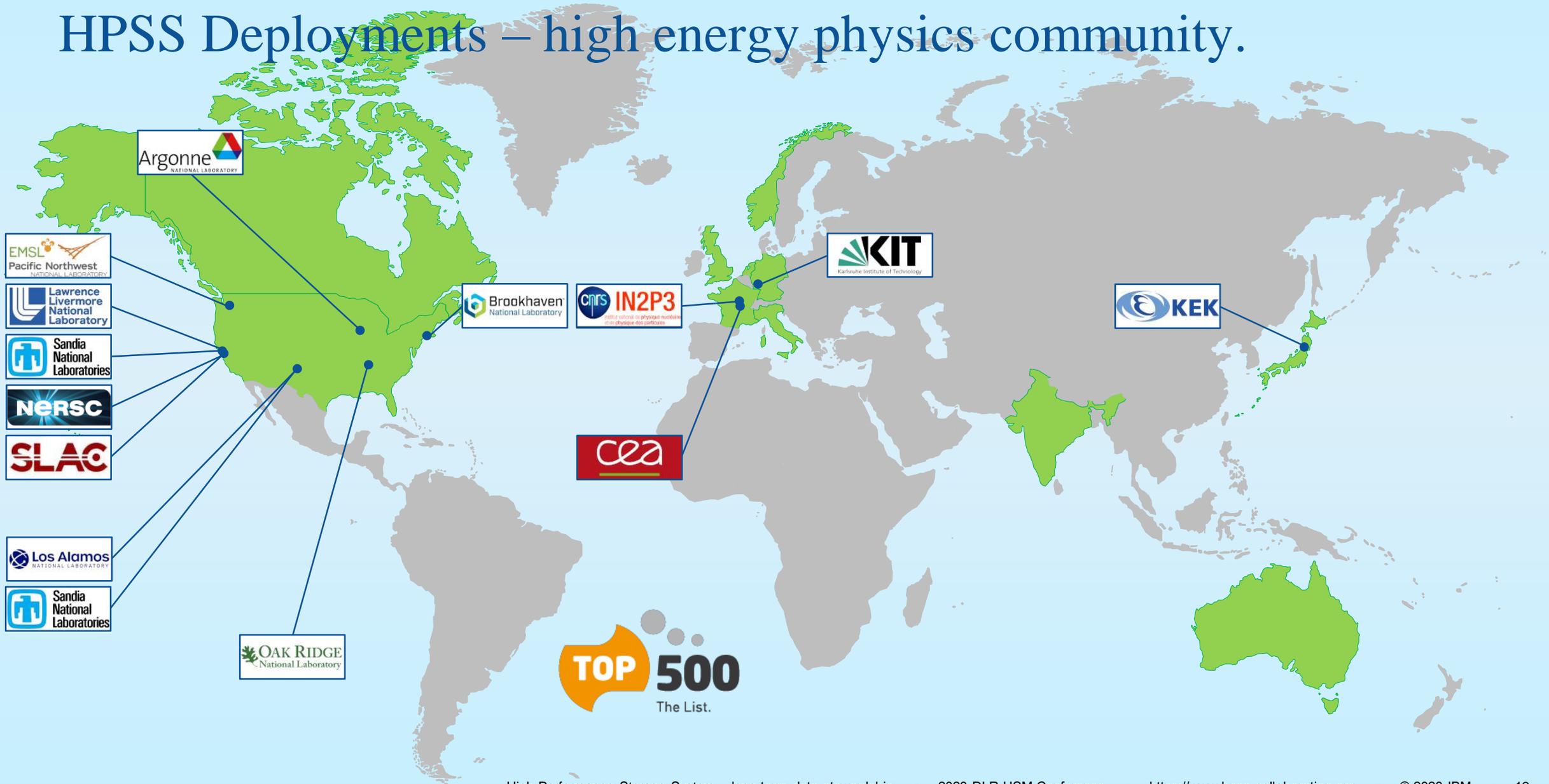
HPSS Deployments – the HPSS collaboration.



HPSS Deployments – other U.S. DoE sites.

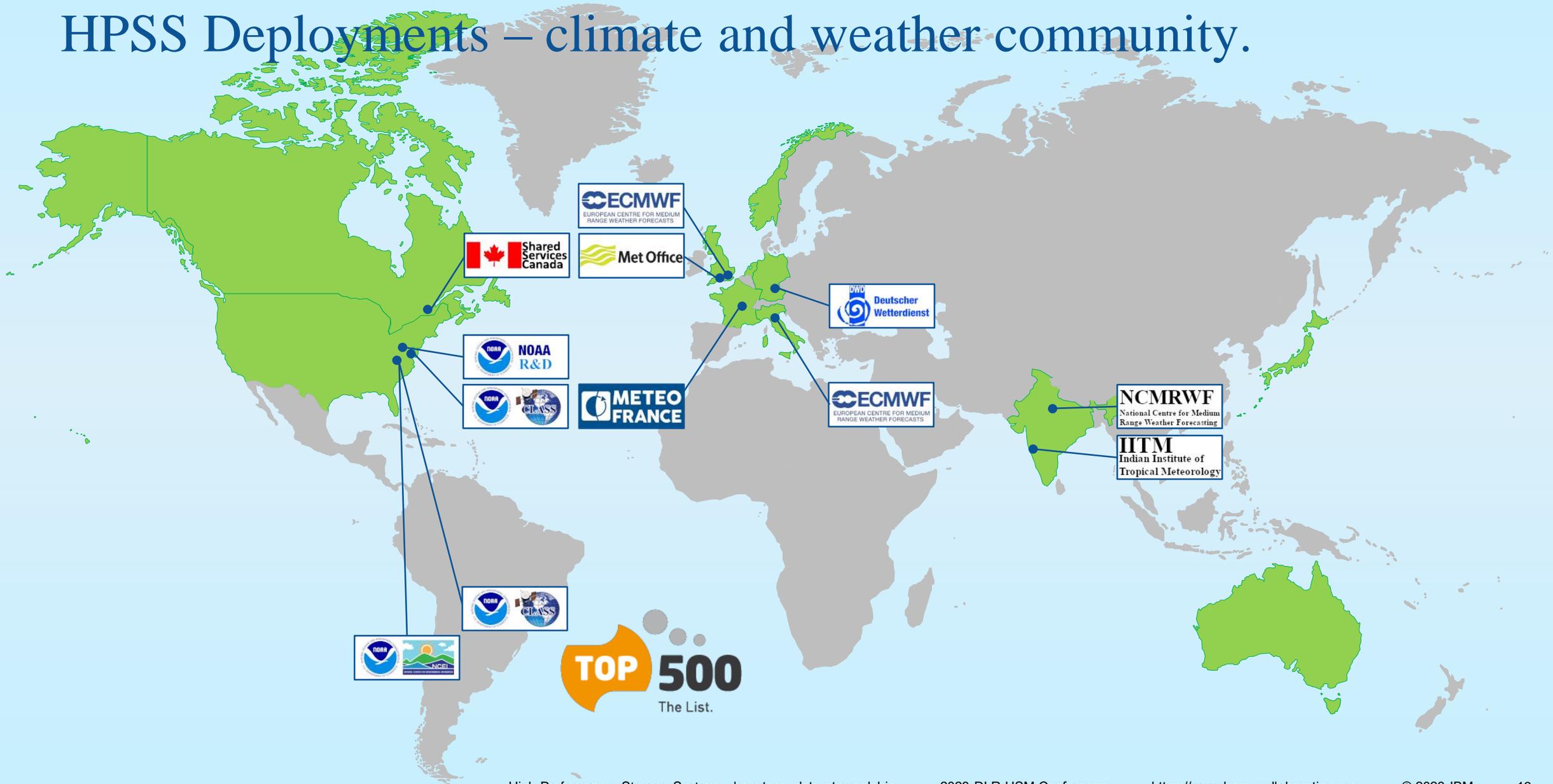


HPSS Deployments – high energy physics community.



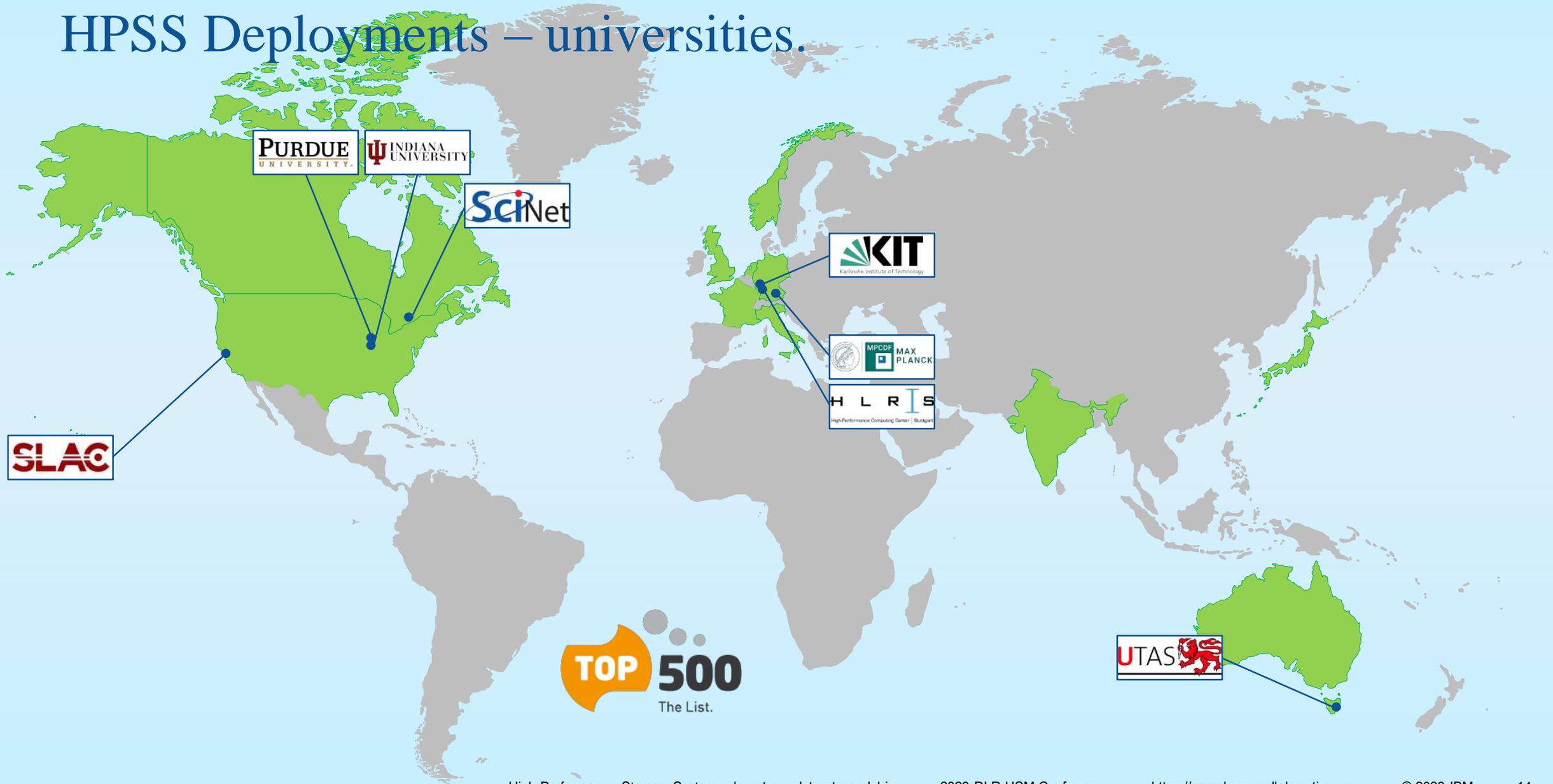


HPSS Deployments – climate and weather community.





HPSS Deployments – universities.



HPSS Deployments – aerospace community.



HPSS Deployments – publicly disclosed.





Large sites use HPSS for obvious reasons!

Sites Names – 1Q2023		PB	M Files	Since
ECMWF	European Centre for Medium-Range Weather Forecasts	778	523	2002
UKMO	United Kingdom Met Office	582	1,349	2009
SSC	Shared Services Canada	515	46	2017
Meteo-France	French Weather and Climate	474	871	2015
NOAA-RD	National Oceanic and Atmospheric Administration Research & Dev.	311	114	2002
LBNL-User	Lawrence Berkeley National Laboratory - User	298	245	1998
MPCDF	Max Planck Computing and Data Facility	242	107	2011
BNL	Brookhaven National Laboratory	220	222	1998
CEA TERA	Commissariat a l'Energie Atomique - Tera Project	219	40	1999
ORNL	Oak Ridge National Laboratory	173	690	1997



HPSS meets the difficult requirements at smaller scales too!

Sites Names – 1Q2023		PB	M Files	Since
IITM	Indian Institute of Tropical Meteorology	19	203	2020
NASA ASDC	NASA Atmospheric Science Data Center	16	60	2018
NASA LaRC	NASA Langley Research Center	13	48	1998
LANL	Los Alamos National Laboratory - Open	12	204	1997
SNL	Sandia National Laboratory - Open	11	189	1996
NCMRWF	India's National Centre for Medium Range Weather Forecasting	10	260	2020
SNL	Sandia National Laboratory - Secure	8	103	1996
NCEI	National Center for Environmental Information	7	125	1998
UTAS	University of Tasmania Research Data Management System	4	66	2020
NB	Nasjonalbiblioteket - National Library of Norway	1	1.5	2022

HPSS – best of breed scaling.

- HPSS scales incrementally.

- Number of files.
- File operations per second.
- Disk and tape bandwidth.
- Disk and tape capacity.
- Client connections.

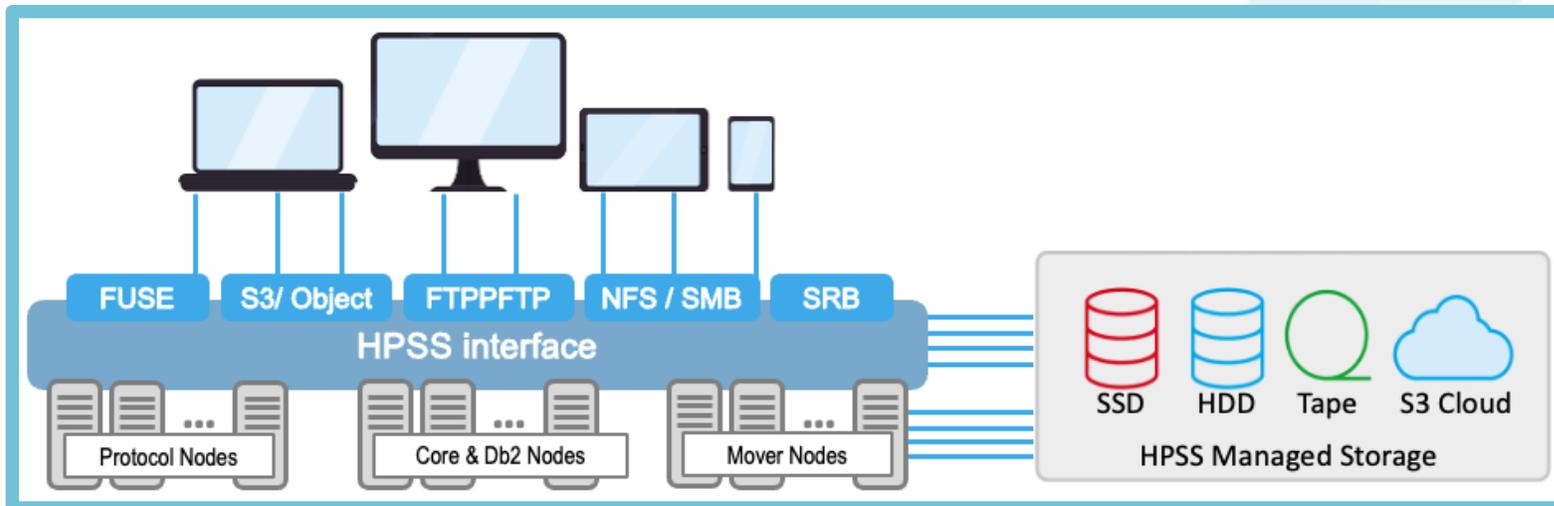
Millions to Billions.

100s File-OPS to 10K File-OPS.

Two movers with tera-scale disk cache & 16 tape drives
to fifty movers with peta-scale disk cache & over 300 tape drives.

A few automated clients to thousands of users.

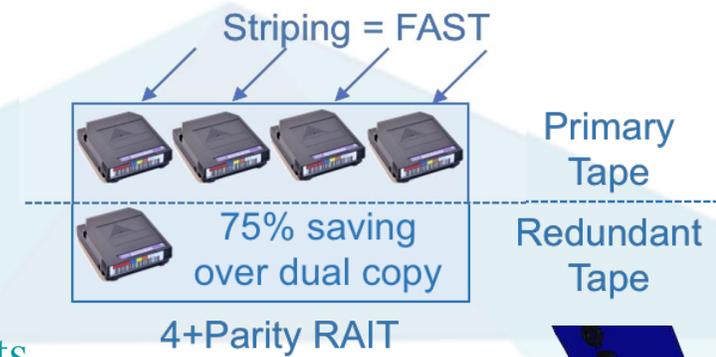
- But... we don't scale our annual support fee.



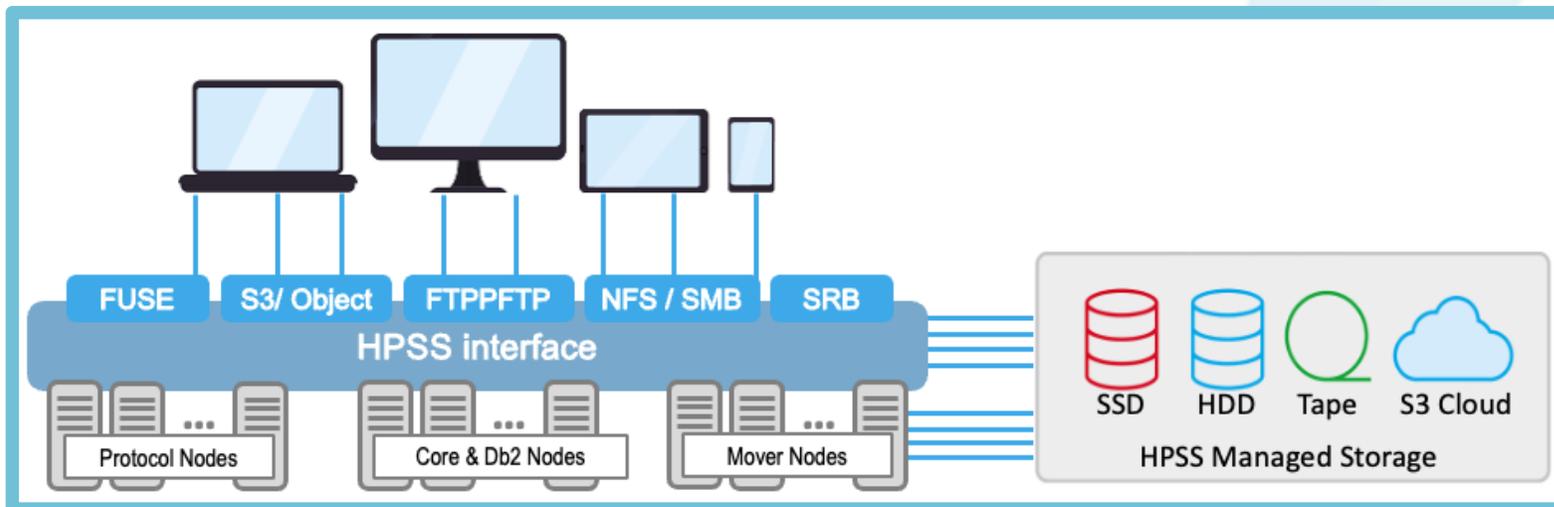
HPSS

HPSS – best of breed high performance features.

- High performance features – we do what most others can't.
 - Extreme-scale single file tape transfers with tape striping.
 - Huge redundant tape cartridge savings with RAIT.
 - Automatic grouping of files and ordered recalls to reduce recall tape mounts.
 - Small file aggregation and full-aggregate recall for improved small transfers.
 - Software support for Recommended Access Order (RAO) to improve recall performance and to extend the life of the tape cartridge by minimizing file-to-file seeks.

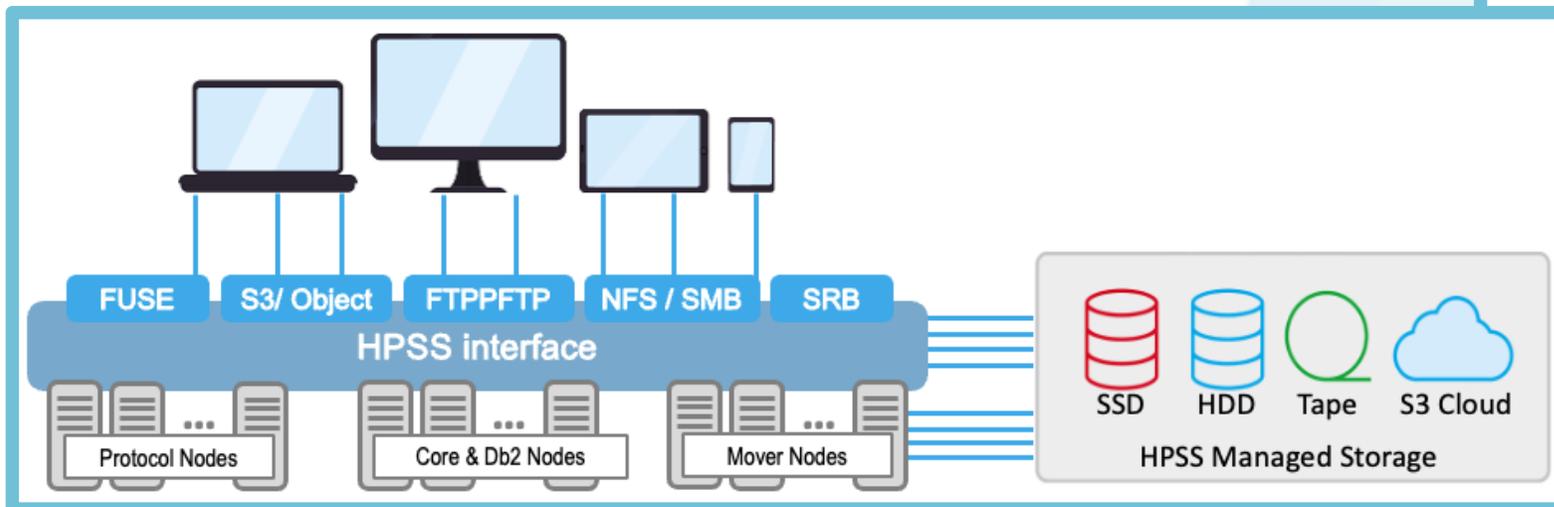


HPSS



HPSS – best of breed battle against data corruption.

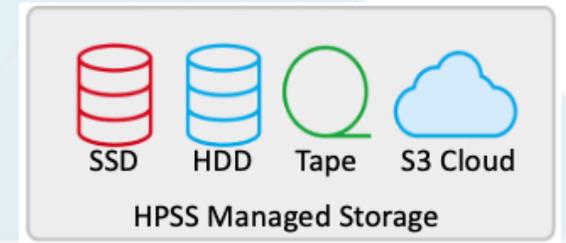
- HPSS looks for silent data corruption in an entirely innovative way!
 - HPSS uses file checksums to ensure data integrity and to verify every file block
 - Logical block protection (LBP) – tape drive feature that requires software integration.
 - HPSS inserts block-level CRCs into the data stream.
 - Our CRC is used by the tape drive.
 - Tape drive read-after-write verifies all data written to tape.
 - “If your file makes it to HPSS tape, it’s 100% true!”



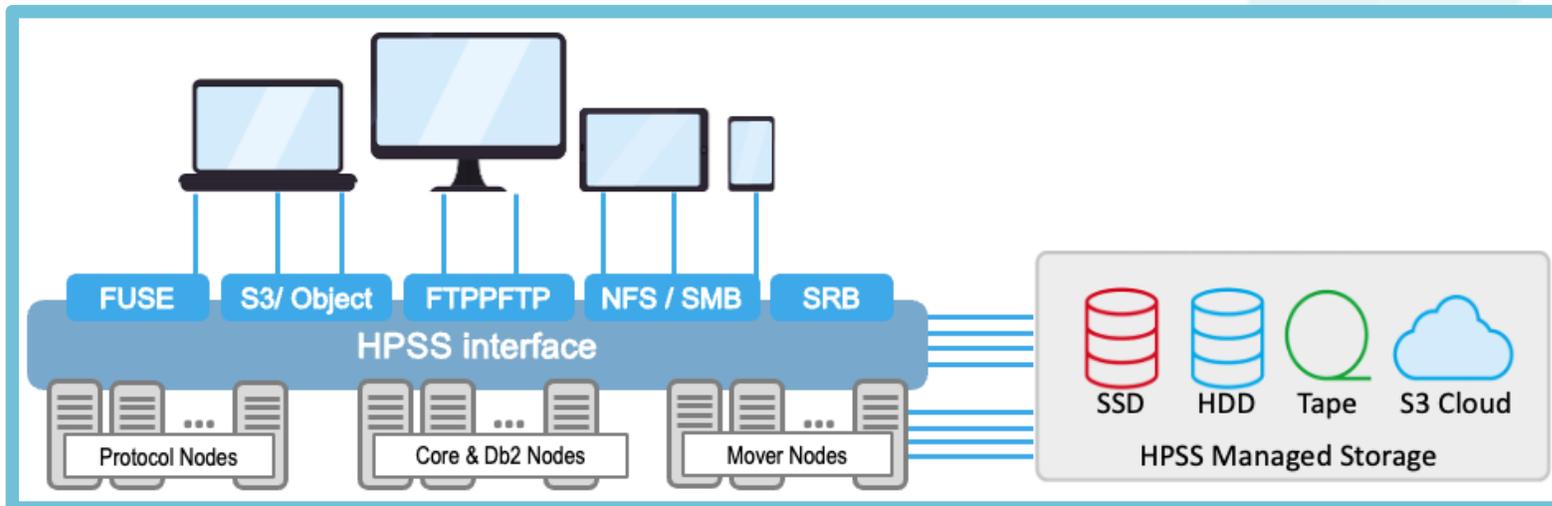
HPSS – best of breed failure recovery and scalability.

- Fast failure recovery.

- HPSS metadata are NOT in an i-node file system.
- All HPSS metadata are in IBM Db2.
- HPSS leverages Db2 partitioning and distribution.
- Db2 transaction logs and backups configured on delivery.
 - Recovered up to the time of failure - Restore backups and roll-forward through the logs in minutes.
- Data and Metadata are always separate – no outage when managed storage is lost.



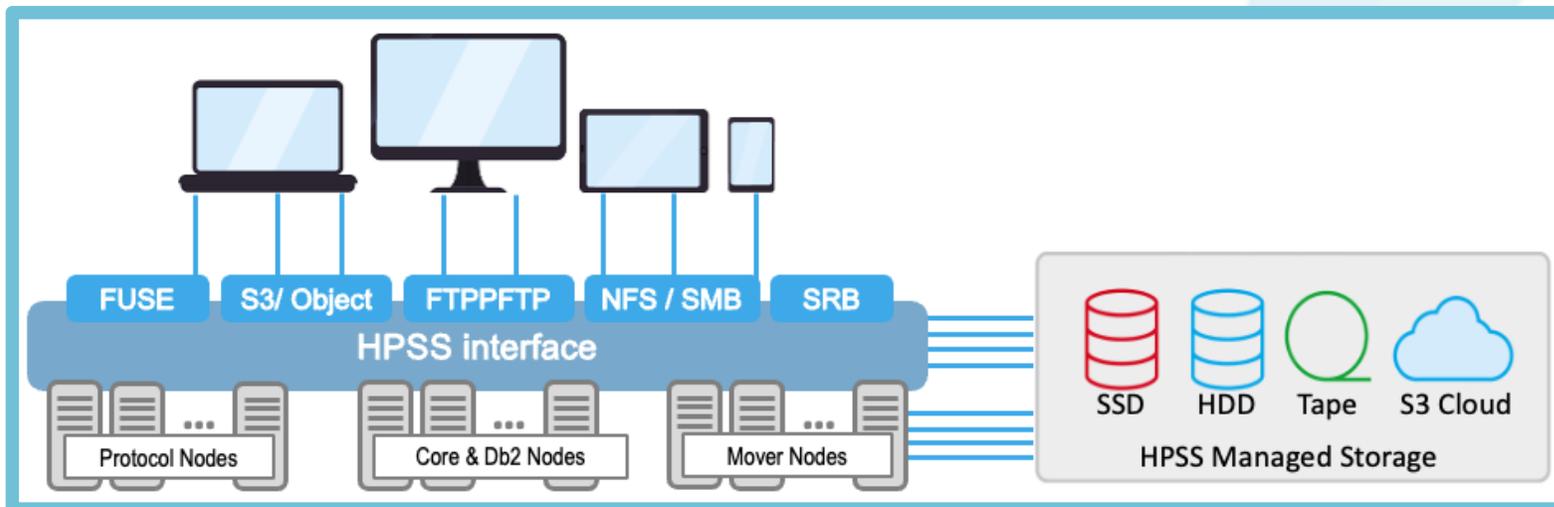
Separation of Metadata and Managed Storage



HPSS

HPSS – best of breed interfaces for flexibility.

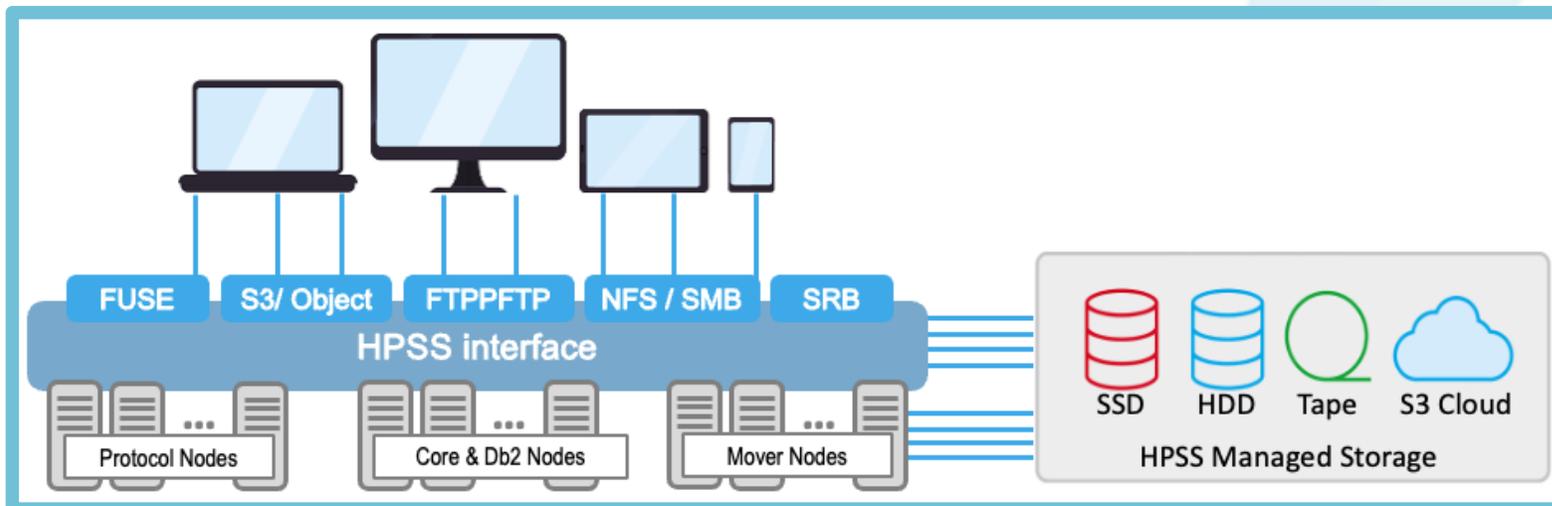
- Lots of ways to store and recall data.
 - HPSS Storage Broker (preservation container).
 - HPSS Tape Archive (HTAR) interface.
 - HPSS S3 (object interface).
 - hpssfs FUSE (file system interface).
 - HPSS parallel FTP daemon.
 - Hierarchical Storage Interface (HSI)
- GHI for Scale – space manage and backups.
- Used within the HPSS community.
 - Globus
 - DSpace
 - NFS
 - Open Samba
 - Bacula
 - Secure FTP
 - iRODS
 - Starfish
 - Lustre RobinHood
 - Lustre CopyTool



HPSS

Convert from Oracle HSM to HPSS

- Move multiple Oracle HSM into a single incrementally scalable HPSS namespace.
- Proven metadata conversion strategy.
 - Our smallest was four Oracle HSM to one HPSS.
 - Our largest was eighteen Oracle HSM to one HPSS.
- Performed in a short down-window.



“Archive” – now what comes to mind with HPSS?



Five Diamondback tape libraries working in parallel



Many thanks!

Jim Gerry

IBM Consulting

jgerry@us.ibm.com

713-256-8516

Let's talk!

HPSS