HPSS @ SC17 - Overview

High Performance Storage System

The value and benefits of the HPSS service offering
We are storage industry thought leaders

• HPSS is a development collaboration that is celebrating our silver anniversary; a partnership between IBM and five DOE labs
  o No single organization has the experience and resources to meet all the challenges represented by the growing imbalance between computing power and data collection capabilities, and storage system I/O, capacity, and functionality

• HPSS is software for long-lived data repositories, for customers that understand the value and benefits of tiered storage
  o Simply stated, not all data needs to be on expensive media, but
  o Access to data on all tiers, and movement of data between tiers MUST be automated and performed in a hardware efficient manner

• HPSS is an IBM service offering
  o Services are best for the stewardship of long lived data

http://www.hpss-collaboration.org
The value of the HPSS service offering

• Annual support fee for 5 PB, 50 PB or 500 PB is the same and remains relatively flat from year to year
  o Greg Lefelar, with Jeskell Systems (an IBM Premier Business Partner), leverages HPSS for tape-requirements that exceed 5 PB

• Delivery services
  o Architect the vendor neutral storage solution with hardware vendors
  o Verify hardware is installed and meets expectations
  o Install and configure HPSS according to delivery milestones
  o Training is accomplished during the delivery process, while classroom training is optional
  o Production readiness review is the final delivery milestone where the keys to a production-ready HPSS system are handed-over

• Personalized support and relationships are cornerstone to the HPSS service offering
Proven technologies for long-term storage

• The industry is re-learning the lessons of days past, HPSS continues to trust in an architecture of scale and reliability

• HPSS metadata are stored in IBM Db2 tables and high speed indexes are leveraged for sorting and locating file details
  o While i-node based solutions typically take hours to scan, suffer from long failure recovery times, and scale by fragmenting the namespace, HPSS does NOT
  o DB2 allows HPSS to quickly recovery up to the point of failure using Db2 logs, Db2 log archive and Db2 backups.
  o Solid state devices, and Db2 partitioning allow HPSS to scale file-operations-per-second, and the number of files stored in a single HPSS
  o Only requiring a couple of terabytes for the active Db2 tables per billion files stored in HPSS

http://www.hpss-collaboration.org
HPSS is ‘Best of Breed for Tape’

• Striping to move large files quickly
• HPSS RAIT (tape stripe with rotating parity) to reduce cost of redundant tape
• Client side aggregation and auto-aggregation for improved small files transfer rates
• Recommended Access Order (RAO) for faster tape recall
• HPSS end-to-end data integrity to identify silent data corruption
• Extreme scale tape mount/dismount logic to maximize the number of tape mounts per hour from each tape library

http://www.hpss-collaboration.org
The HPSS big picture

Massively scalable global HPSS namespace enabled by DB2

RHEL Core Server & Mover computers
Intel  Power

Extreme-scale high-performance automated HSM
Disk  Tape

Block or Filesystem Disk Tiers
Hardware Vendor Neutral

Enterprise  LTO Tape
IBM  Oracle  Spectra Logic

http://www.hpss-collaboration.org

HPSS @ SC17– Overview
Scaling HPSS architecture

**EXTREMELY SCALABLE**

- Adding storage units for more disk cache capacity and bandwidth
- Adding libraries for higher mount rates and higher capacity
- Adding tape drives for higher bandwidth, file recalls, repacks and validation

![Diagram of HPSS architecture](http://www.hpss-collaboration.org)

- Adding interface protocol servers for more client connections
- Adding movers for increased tape drives and disk cache storage units
- Dedicated Tape Movers with 32-cores support 16 tape drives

**HPSS @ SC17 – Overview**
Publically disclose HPSS customers
Publically disclose HPSS customers

<table>
<thead>
<tr>
<th>Sites</th>
<th>10^{15} Bytes</th>
<th>10^6 Files</th>
</tr>
</thead>
<tbody>
<tr>
<td>(ECMWF) European Centre for Medium-Range Weather Forecasts</td>
<td>335.88</td>
<td>317.22</td>
</tr>
<tr>
<td>(UKMO) United Kingdom Met Office</td>
<td>199.56</td>
<td>195.72</td>
</tr>
<tr>
<td>(NOAA-RD) National Oceanic and Atmospheric Administration Research &amp; Development</td>
<td>120.90</td>
<td>89.24</td>
</tr>
<tr>
<td>(BNL) Brookhaven National Laboratory</td>
<td>119.46</td>
<td>134.45</td>
</tr>
<tr>
<td>(LBNL-User) Lawrence Berkley National Laboratory - User</td>
<td>113.01</td>
<td>220.41</td>
</tr>
<tr>
<td>(Meteo-France) Meteo France</td>
<td>92.85</td>
<td>355.73</td>
</tr>
<tr>
<td>(CEA TERA) Commissariat a l’Energie Atomique - GENO</td>
<td>83.54</td>
<td>17.79</td>
</tr>
<tr>
<td>(NCAR) National Center for Atmospheric Research</td>
<td>83.05</td>
<td>272.85</td>
</tr>
<tr>
<td>(MPCDF) Max Planck</td>
<td>78.13</td>
<td>168.42</td>
</tr>
<tr>
<td>(ORNL) Oak Ridge National Laboratory</td>
<td>74.66</td>
<td>77.97</td>
</tr>
<tr>
<td>(LANL-Secure) Los Alamos National Laboratory - Secure</td>
<td>73.55</td>
<td>716.06</td>
</tr>
<tr>
<td>(LLNL-Secure) Lawrence Livermore National Laboratory - Secure</td>
<td>69.40</td>
<td>913.45</td>
</tr>
<tr>
<td>(DKRZ) Deutsches Klimarechenzentrum</td>
<td>63.69</td>
<td>18.77</td>
</tr>
<tr>
<td>(IN2P3) Institut National de Physique Nucleaire et de Physique des Particules</td>
<td>52.56</td>
<td>69.93</td>
</tr>
</tbody>
</table>

Capacity Leader
HPSS Treefrog Demo @ SC17

- Treefrog prototype browser interface
- Storing and retrieving datasets
  - 1,500 1MB files in the demo dataset
- Treefrog error correction encoding
  - Two data + one parity tape
- LTO-6 tape technology with a 160 MB/s native transfer rate
- Treefrog demonstrating beyond 128 MB/s per tape drive storing and retrieving 1MB files
- Dataset transfers beyond 256 MB/s
Thank you!

Jim Gerry
jgerry@us.ibm.com