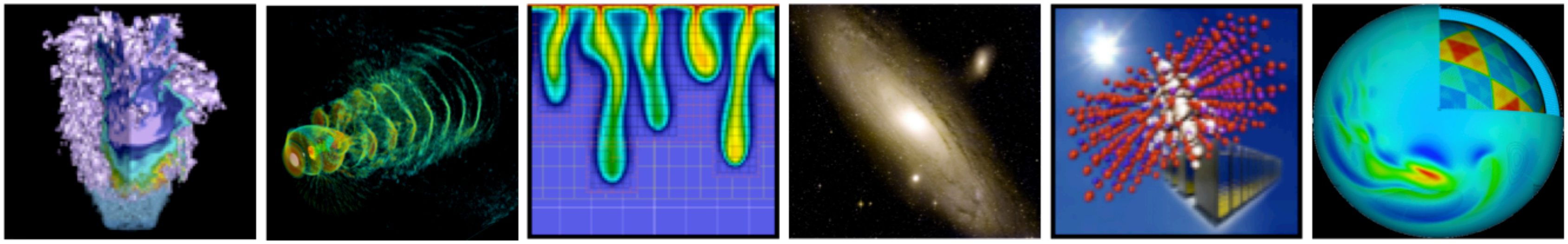


Oracle OpenWorld 2011: Digital Archiving and Preservation in Government Departments and Agencies

Jason Hick

jhick@lbl.gov

NERSC LBNL

http://www.nersc.gov/nusers/systems/HPSS/

October 6, 2011

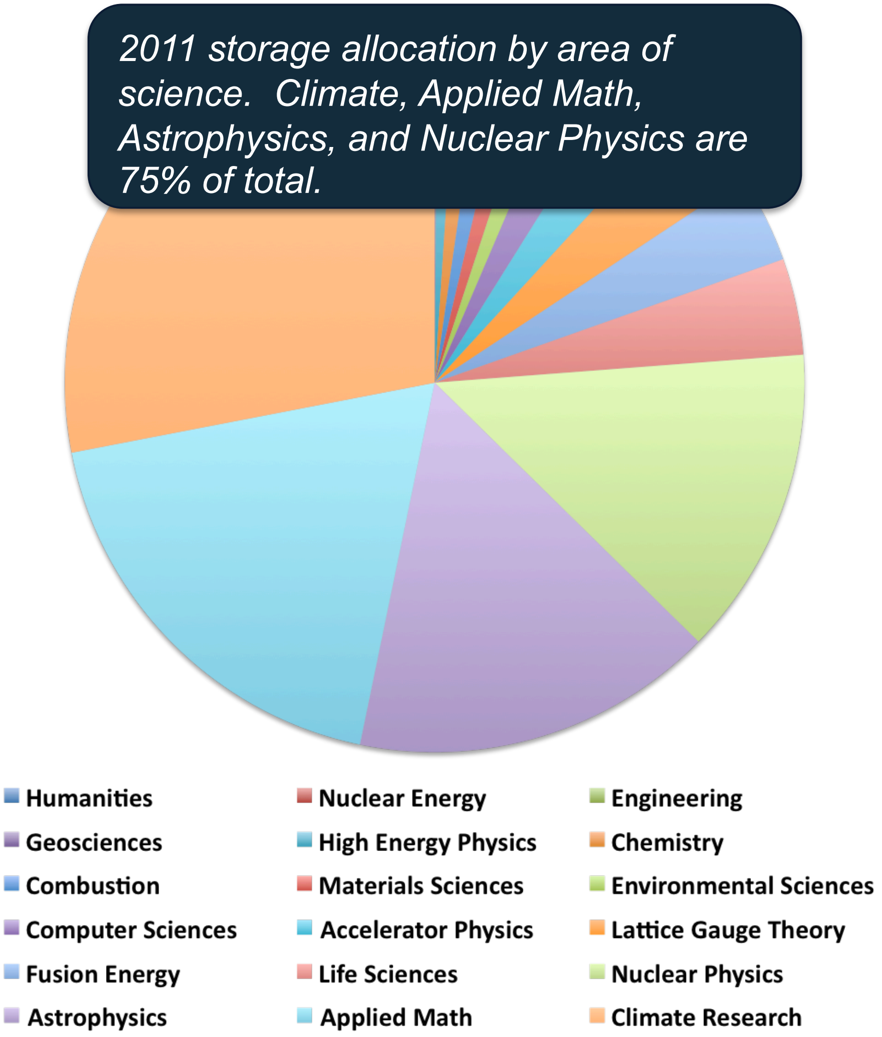
U.S. DEPARTMENT OF ENERGY | Office of Science

NERSC National Energy Research Scientific Computing Center

BERKELEY LAB Lawrence Berkeley National Laboratory

The Production Facility for DOE Office of Science

- **Operated by the University of California for the U.S. DOE**
- **NERSC serves a large population**
  - Approximately 4000 users, 400 projects, 500 codes
  - Focus on “unique” resources
    - High-end computing systems
    - High-end storage systems
      - Large shared GPFS (a.k.a. NGF)
      - Large archive (a.k.a. HPSS)
    - Interface to high speed networking
      - ESnet border soon to be 100Gb
- **Our mission is to accelerate the pace of discovery by providing high performance computing, data, and communication services to the DOE Office of Science community.**



2011 storage allocation by area of science. Climate, Applied Math, Astrophysics, and Nuclear Physics are 75% of total.

Focused on Data Needs

- **We present efficient center-wide storage solutions**
  - NGF, a centralized center-wide file system, aids in minimizing multiple online copies of data and reduces extraneous data movement between systems.
  - HPSS enables exponential user data growth without exponential impact to the facility, and provides long-term storage to our users.
- **Partnering with ANL and ESnet to advance HPC network capabilities**
  - Magellan ANL and NERSC cloud research infrastructure.
  - Advanced Networking Initiative with ESnet (100Gb Ethernet).
  - Leadership in inter-site data transfers (Data Transfer Working Group: ORNL, ANL, NERSC, LANL, and ESnet).
- **We are a distribution point for scientific data**
  - Sudbury Nutrino Observatory (SNO) archive. We retain about 70TBs of detector data that provides revolutionary insight into the property of neutrinos and the core of the sun.
  - The Gauge Connection. A web gateway to an archive for lattice quantum chromodynamics (QCD). A repository of gauge configurations for understanding the behavior of quarks and gluons.
  - DeepSky. A web interface to astronomical data enabling collaborative discoveries of supernovae. Over 600 discovered since 2010, and the closest in 25 years discovered hours after it exploded (Sep 2011)
  - Earth Systems Grid (ESG) Gateway. Various climate data sets.

The Storage Systems Group

- Wayne Hurlbert and Nick Balthaser: HPSS system analysts
- Damian Hazen and Mike Welcome: HPSS developers
- Matt Andrews: NGF backup developer
- Will Baird: Data transfer system analyst
- Rei Lee and Greg Butler: NGF system analysts

Storage Services Offered

- **Center-wide online storage to minimize data movement and duplication.**
- **Archival storage for long-term data retention.**
- **Science gateways for presenting and sharing data repositories on the web.**
- **Data transfer solutions to aid in inter-site data movement.**
  - Globus Online (http://www.globusonline.org)
  - GridFTP
  - bbcp

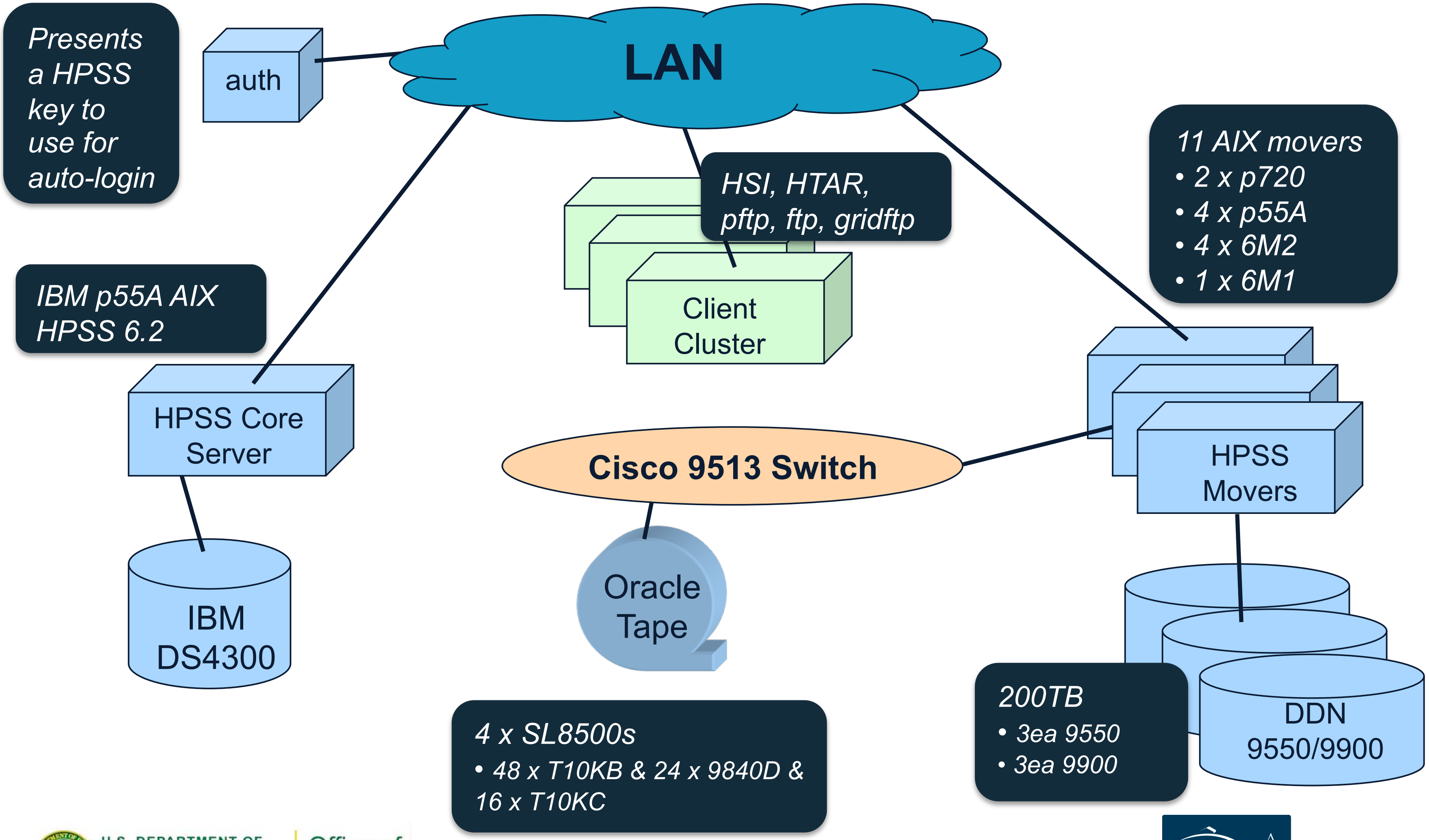
Center-wide File Systems

- **/project is for sharing and long-term residence of data on all NERSC computational systems.**
  - 4% monthly growth, ~50% growth per year
  - Not purged, quota enforced (4TB default per project), backed up daily
  - Serves 200 projects over FC4/8
  - 1.6PB total capacity
  - ~5TB average daily IO
- **/global/homes provides a common login environment for users across systems.**
  - 7% monthly growth, 85% growth per year
  - Not purged but archived, quota enforced (40GB per user), backed up daily
  - Serves 4000 users, 400 per day over Ethernet
  - 50TB total capacity
  - 100’s of GBs average daily IO
- **/global/scratch provides high bandwidth and capacity data across systems.**
  - Purged, quota enforced (20TB per user), not backed up
  - Serves 4000 users over FC8
  - 1PB total capacity

Archival Storage

- **User HPSS (~12 PB as of 9/30/2011)**
  - Single transfers 1GB/sec read/write
  - Aggregate bandwidth 4+GB/sec
  - Average daily IO of 20TB, with peak at 40TB
  - 200TB disk cache
  - 24 9840D, 48 T10KB, 16 T10KC tape drives
  - Largest file: 5.5TB
  - Oldest file: Jan 1976
- **Backup HPSS (~13 PB as of 9/30/2011)**
  - Single transfers 1GB/sec read/write
  - Aggregrate bandwidth 3+GB/sec
  - Average daily IO of 10TB, with peak at 130TB
  - 40TB disk cache
  - 8 9840D and 18 T10KB tape drives
  - Largest file: 3.5TB
  - Oldest file: May 1995

User HPSS Configuration



Presents a HPSS key to use for auto-login

auth

LAN

HSI, HTAR, pftp, ftp, gridftp

Client Cluster

11 AIX movers
- 2 x p720
- 4 x p55A
- 4 x 6M2
- 1 x 6M1

IBM p55A AIX HPSS 6.2

HPSS Core Server

Cisco 9513 Switch

HPSS Movers

IBM DS4300

Oracle Tape

DDN 9550/9900

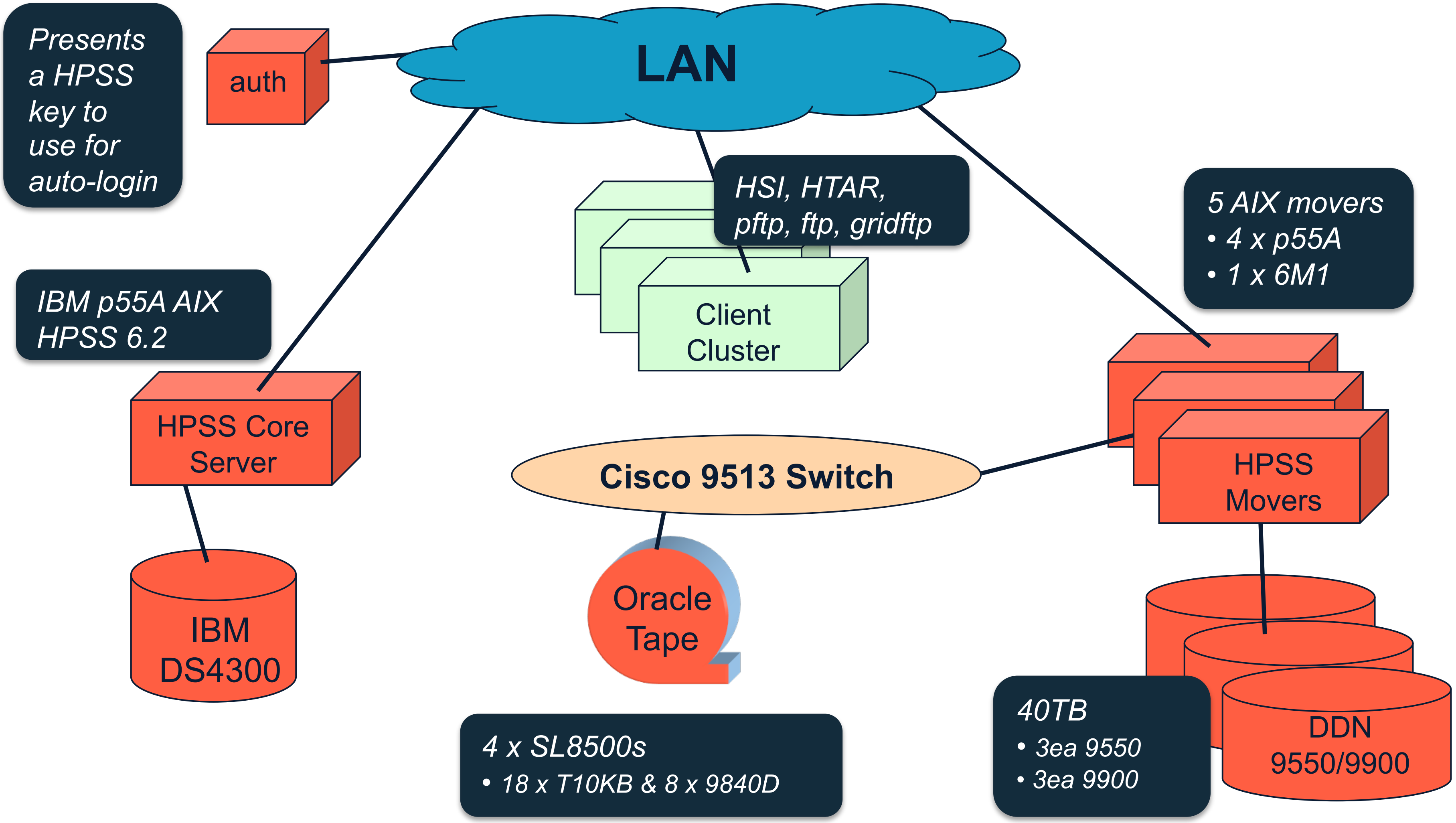
200TB
- 3ea 9550
- 3ea 9900

4 x SL8500s
- 48 x T10KB & 24 x 9840D & 16 x T10KC

U.S. DEPARTMENT OF ENERGY | Office of Science

BERKELEY LAB Lawrence Berkeley National Laboratory

Backup HPSS Configuration



Presents a HPSS key to use for auto-login

IBM p55A AIX
HPSS 6.2

HSI, HTAR, pftp, ftp, gridftp

5 AIX movers
- 4 x p55A
- 1 x 6M1

4 x SL8500s
- 18 x T10KB & 8 x 9840D

40TB
- 3ea 9550
- 3ea 9900

Why we use disk

- **Analysis of data**
  - File systems, databases, high data and metadata rates required
- **Distribution of frequently requested data sets**
  - Disk is spinning anyways, best to utilize available bandwidth where possible
- **Computational system interaction**
  - File systems are expected
- **Random I/O**
  - Random access is possible

Why we use tape

- **Exponential growth with a reasonable budget**
  - Capacity of tape doubles about every two years and no end in sight soon (at least 10 more years of this with today's tape head technology)
- **Has capacity/growth characteristics that make exabyte archives feasible (Exascale plans)**
- **Long term preservation aspects**
  - decadal lifetime vs. 3-5 year lifetime for disk
- **Facility operational efficiency**
  - Power, cooling orders of magnitude lower than disk
- **Provides separation and redundancy of storage technologies (tape & disk)**
  - Ideal for PB-sized backups

Large Tape Users Group

- **An IOUG special interest group focused on the StorageTek tape technology products**
  - Share user experiences
  - Provide user feedback and requirements to Oracle
- **Membership requirements**
  - Own or manage 10,000 slots, or, have 1PB of data stored in one or more Oracle STK tape library
- **If you meet the requirements above, we encourage you to join:**
  - Website: http://ltug.oracle.ioug.org/
- **Annual user conference in Broomfield, CO**
  - LTUG 2012, April 23-26