

UCSD-CAIDA Hosting Infrastructure Reports

UCSD-CAIDA 2012

1. *Description*

- To support the requirements that come with the roles of PREDICT Data Host (DH) and Data Provider (DP), the CAIDA group at UCSD maintains numerous computers mounted in the machine room at the San Diego Supercomputer Center. Our system administrators have designed, configured and deployed these hosts to provide high availability for data collection, indexing, curation, and distribution to researchers approved via the PREDICT vetting mechanism.
- As a general strategy, we have chosen to deploy several hosts with moderately large (20-40TB), locally attached disk systems that make use of the ZFS file systems. These configurations optimize for cost of storage and availability for data consumers. We also run several systems that act as web servers for hosting project description pages as well as for data distribution to vetted account holders. We have begun making use of FreeBSD jails servers that mount exported file systems from the backend data servers.
- Separately, as an experiment to compare cost of ownership, reliability, and performance, we also purchase cloud disk storage from our host department, SDSC. We use this storage for archival of datasets and plan to also utilize the services for distribution of data.
- Additionally, we make use of an Energy Research Computing Allocations Process (ERCAP) Allocation at the National Energy Research Scientific Computing Center (NERSC) facility, a division of the Lawrence Berkeley National Laboratory located in Berkeley, California. SDSC enjoys high bandwidth connectivity (10GB) with the NERSC.GOV domain allowing to conduct regular file transfers for archival of historical data.

2. *System Inventory*

- Data Server
 - o OS: FreeBSD 8.2
 - o CPUs: 1 x 4 core Intel(R) Xeon(R) CPU E5620 @ 2.40GHz
 - o Memory: 6GB RAM
 - o Storage: 48 TB raw disk (38 TB after RAID 6 and zfs overhead)
 - o Description: This machine is our main data server.
- Web Server
 - o OS: FreeBSD 8.2 (jail server)
 - o CPUs: 2 x 6 core Intel(R) Xeon(R) CPU X5670 @ 2.93GHz
 - o Software: Apache
 - o Memory: 48GB RAM
 - o Storage: minimum requirements for jailed operating system and mounted file systems.
 - o Description: The FreeBSD jail server that mounts the exported file systems. We run two redundant jail servers so that one can take over if the other fails.
- Web Server
 - o OS: FreeBSD 8.2 (jail server)
 - o CPUs: 2 x 6 core Intel(R) Xeon(R) CPU X5670 @ 2.93GHz
 - o Software: Apache

- Memory: 48GB RAM
- Storage: minimum requirements for jailed operating system and mounted file systems.
- Description: This machine is the backup jail server.
- Web Server
 - OS: FreeBSD 7.3
 - CPUs: 2 x 4 core Intel(R) Xeon(R) CPU X5667 @ 3.07GHz
 - Software: Apache
 - Memory: 32GB RAM
 - Storage: 12 TB of raw disk (~9TB useable after RAID 5 and filesystem overhead)
 - Description: This host is the main data server for the Archipelago measurement infrastructure.
- Data Server
 - OS: FreeBSD 8.3
 - CPUs: 2 x 6 core Intel(R) X5675 @ 3.07GHz
 - Memory: 192GB RAM
 - Storage: 102 TB raw disk (74 TB after RAID 6 and zfs overhead)
 - Description: This host acts both as the primary data server and the primary analysis machine for the UCSD Network Telescope data.
- Web Server
 - OS: FreeBSD 6.2
 - CPU: 1 x 2 core Intel(R) Pentium(R) D CPU 3.00GHz
 - Software: Apache
 - Memory: 1GB RAM
 - Storage: 152GB local disk
 - Description: This server mounts all topology data over NFS
- Data and Compute Server
 - OS: FreeBSD 8.3
 - CPUs: 1 x 4 core Intel(R) Xeon(R) CPU W3530 @ 2.80GHz
 - Memory: 6GB RAM
 - Storage: 48 TB raw disk (38 TB after RAID 6 and zfs overhead)
 - Description: This server hosts the UCSD Network Telescope data and acts as the main compute server for near real-time telescope data.
- Web Server
 - OS: FreeBSD 5.4
 - CPUs: 1 x 2 core Intel(R) Pentium(R) D CPU 3.00GHz
 - Software: Apache
 - Memory: 3GB RAM
 - Storage: 73GB local disk
 - Description: This server supports the web infrastructure that serves the dataset description pages, forms, and project web pages.
- Data Archival and Storage
 - OS: Linux (Rocks Clustering tool kit)
 - Software: OpenStack (Swift)
 - Storage: 45 TB

- Description: We purchase cloud disk storage space from SDSC for storing our system backups and archiving PREDICT data. 20 TB of this space originally came as a transitional allocation from SDSC during graduate decommissioning of HPSS and SamQFS tape facilities. In the coming year, we will be charged for the entire 45 TB.
- NERSC HPSS Tape Archive Allocation
 - OS: AIX
 - Software: HPSS Tape Services
 - Storage: 150 TB
 - Description: We enjoy a 150 TB allocation of HPSS tape resources at the NERSC facility where we archive our historical UCSD Network Telescope (darknet) data. As of the end of 2012, we have used approximately 105TB of this allocation.

UCSD-CAIDA 2013

1. *Description*

- To support the requirements that come with the roles of PREDICT Data Host (DH) and Data Provider (DP), the CAIDA group at UCSD maintains numerous computers hosted in the machine room at the San Diego Supercomputer Center. Our system administrators have designed, configured and deployed these hosts to provide high availability for data collection, indexing, curation, and distribution to researchers vetted via the PREDICT portal.
- As a general strategy, we have chosen to deploy several hosts with moderately large (20-40TB) locally attached disk systems that make use of the ZFS file systems. These configurations optimize for cost of storage and availability for data consumers. We also run several systems that act as web servers hosting project description pages and distributing data to vetted account holders. We use FreeBSD jails servers that mount exported file systems from the backend data servers.
- As an experiment to compare cost of ownership, reliability, and performance, we began using cloud disk storage from our host department, SDSC for archival of scientific datasets.
- Finally, we make use of an Energy Research Computing Allocations Process (ERCAP) Allocation at the National Energy Research Scientific Computing Center (NERSC) facility, a division of the Lawrence Berkeley National Laboratory located in Berkeley, California. SDSC has high bandwidth connectivity (10 GB) with the NERSC.GOV domain allowing to conduct regular file transfers for archival of historical data.

2. *System Inventory*

- Data Server
 - OS: FreeBSD 8.2
 - CPUs: 1 x 4 core Intel(R) Xeon(R) CPU E5620 @ 2.40 GHz
 - Memory: 6 GB RAM
 - Storage: 48 TB raw disk (38 TB after RAID 6 and ZFS overhead)
 - Description: This machine is our main data server.
- Web Server
 - OS: FreeBSD 9.2 (jail server)
 - CPUs: 2 x 6 core Intel(R) Xeon(R) CPU X5670 @ 2.93 GHz
 - Software: Apache

- Memory: 48 GB RAM
- Storage: minimum requirements for jailed operating system and mounted file systems.
- Description: The FreeBSD jail server that mounts the exported file systems. We run two redundant jail servers so that one can take over if the other fails.
- Web Server
 - OS: FreeBSD 8.2 (jail server)
 - CPUs: 2 x 6 core Intel(R) Xeon(R) CPU X5670 @ 2.93 GHz
 - Software: Apache
 - Memory: 48 GB RAM
 - Storage: minimum requirements for jailed operating system and mounted file systems.
 - Description: This machine is the backup jail server.
- Web Server
 - OS: Ubuntu 14.04
 - CPUs: 2 x 4 core Intel(R) Xeon(R) CPU X5667 @ 3.07 GHz
 - Memory: 32 GB RAM
 - Storage: 12 TB of raw disk (~9 TB useable after RAID 5 and file system overhead)
 - Description: This host is the main data server for the Archipelago measurement infrastructure.
- Data Server
 - OS: FreeBSD 10
 - CPUs: 2 x 6 core Intel(R) X5675 @ 3.07 GHz
 - Memory: 192 GB RAM
 - Storage: 102 TB raw disk (74 TB after RAID 6 and ZFS overhead)
 - Description: This host acts both as the primary data server and as the primary analysis machine for the UCSD Network Telescope data.
- Data and Compute Server
 - OS: FreeBSD 10
 - CPUs: 1 x 4 core Intel® Xeon® CPU W3530 2.8 GHz
 - Memory: 6 GB RAM
 - Storage: 48 TB raw disk (38 TB after RAID 6 and zfs overhead)
 - Description: This server hosts the UCSD Network Telescope data and acts as the main compute server for near real time telescope data.
- Web Server
 - OS: FreeBSD 5.4
 - CPUs: 1 x 2 core Intel(R) Pentium(R) D CPU 3.00 GHz
 - Software: Apache
 - Memory: 3 GB RAM
 - Storage: 73 GB local disk
 - Description: This server supports the web infrastructure that serves the dataset description pages, forms, and project web.
- Data Archival and Storage
 - OS: Linux (Rocks Clustering tool kit)
 - Software: OpenStack (Swift)
 - Storage: 45 TB

- Description: We are paying for cloud disk storage space from SDSC for storing our system backups and archiving PREDICT data.
- NERSC HPSS Tape Archive Allocation
 - OS: AIX
 - Software: HPSS Tape Services
 - Storage: 200 TB
 - Description: We archive our historical UCSD Network Telescope (darknet) data using HPSS tape resources at the NERSC facility. As of July 2013, we used approximately 60% of our allocated resources for this calendar year. We plan to continue submitting annual applications for more space as needed.

UCSD-CAIDA 2014

1. Description

- To support the requirements that come with the roles of PREDICT Data Host (DH) and Data Provider (DP), the CAIDA group at UCSD maintains numerous computers hosted in the machine room at the San Diego Supercomputer Center. Our system administrators have designed, configured and deployed these hosts to provide high availability for data collection, indexing, curation, and distribution to researchers vetted via the PREDICT portal.
- As a general strategy, we have chosen to deploy several hosts with moderately large (20-40TB) locally attached disk systems that make use of the ZFS file systems. These configurations optimize for cost of storage and availability for data consumers. We also run several systems that act as web servers hosting project description pages and distributing data to vetted account holders. We use FreeBSD jails servers that mount exported file systems from the backend data servers.
- As an experiment to compare cost of ownership, reliability, and performance, we have been using cloud disk storage from our host department, SDSC for archival of scientific datasets.
- Finally, we make use of an Energy Research Computing Allocations Process (ERCAP) Allocation at the National Energy Research Scientific Computing Center (NERSC) facility, a division of the Lawrence Berkeley National Laboratory located in Berkeley, California. SDSC has high bandwidth connectivity (10 GB) with the NERSC.GOV domain allowing to conduct regular file transfers for archival of historical data.

2. System Inventory

- Data Server
 - OS: FreeBSD 8.2
 - CPUs: 1 x 4 core Intel(R) Xeon(R) CPU E5620 @ 2.40 GHz
 - Memory: 6 GB RAM
 - Storage: 48 TB raw disk (38 TB after RAID 6 and ZFS overhead)
 - Description: This machine is our main data server.
- Web Server
 - OS: FreeBSD 9.2 (jail server)
 - CPUs: 2 x 6 core Intel(R) Xeon(R) CPU X5670 @ 2.93 GHz
 - Software: Apache

- Memory: 48 GB RAM
- Storage: minimum requirements for jailed operating system and mounted file systems.
- Description: The FreeBSD jail server that mounts the exported file. We run two redundant jail servers so that one can take over if the other fails.
- Web Server
 - OS: FreeBSD 8.2 (jail server)
 - CPUs: 2 x 6 core Intel(R) Xeon(R) CPU X5670 @ 2.93 GHz
 - Software: Apache
 - Memory: 48 GB RAM
 - Storage: minimum requirements for jailed operating system and mounted file systems.
 - Description: This machine is the backup jail server.
- Web Server
 - OS: Ubuntu 14.04
 - CPUs: 2 x 4 core Intel(R) Xeon(R) CPU X5667 @ 3.07 GHz
 - Memory: 32 GB RAM
 - Storage: 12 TB of raw disk (~9 TB useable after RAID 5 and file system overhead)
 - Description: This host is the main data server for the Archipelago measurement infrastructure.
- Data Server
 - OS: FreeBSD 10
 - CPUs: 2 x 6 core Intel(R) X5675 @ 3.07 GHz
 - Memory: 192 GB RAM
 - Storage: 102 TB raw disk (74 TB after RAID 6 and ZFS overhead)
 - Description: This host acts both as the primary data server and as the primary analysis machine for the UCSD Network Telescope data.
- Data and Compute Server
 - OS: FreeBSD 10
 - CPUs: 1 x 6 core Intel E5-1650 3.5 GHz
 - Memory: 32 GB RAM
 - Storage: 72 TB raw disk (43 TB after RAID z3 and 2 hot spares)
 - Description: This server is used for CAIDA general backups. We are migrating backups of scientific data to this machine as well.
- Web Server
 - OS: FreeBSD 5.4
 - CPUs: 1 x 2 core Intel(R) Pentium(R) D CPU 3.00 GHz
 - Software: Apache
 - Memory: 3 GB RAM
 - Storage: 73 GB local disk
 - Description: This server supports the web infrastructure that serves the dataset description pages, forms, and project web pages. We plan to switch to a jail server system next year.
- Data Archival and Storage
 - OS: Linux (Rocks Clustering tool kit)
 - Software: OpenStack (Swift)

- Storage: 25.5 TB
- Description: We are moving all data currently stored in this cloud system to local CAIDA storage.
- NERSC HPSS Tape Archive Allocation
 - OS: AIX
 - Software: HPSS Tape Services
 - Description: We archive 230 TB of our historical UCSD Network Telescope (darknet) data using HPSS tape resources at the NERSC facility. As of July 2014, we used approximately 48% of our allocated resources for this calendar year.

UCSD-CAIDA 2015

1. *Description*

- To support the requirements that come with the roles of IMPACT Data Host (DH) and Data Provider (DP), the CAIDA group at UCSD maintains numerous computers hosted in the machine room at the San Diego Supercomputer Center. Our system administrators have designed, configured and deployed these hosts to provide high availability for data collection, indexing, curation, and distribution to researchers vetted via the IMPACT portal.
- As a general strategy, we have chosen to deploy several hosts with moderately large (20-40TB) locally attached disk systems that make use of the ZFS file systems. These configurations optimize for cost of storage and availability for data consumers. We also run several systems that act as web servers hosting project description pages and distributing data to vetted account holders. We use FreeBSD jails servers that mount exported file systems from the backend data servers.
- This year we inherited two nodes from the recently decommissioned SDSC Trestles Supercomputer that we repurposed as data and compute servers in support of our new topology query system as well as for conducting MIDAR runs to produce our flagship Internet Topology Data Kit (ITDK) datasets.
- In support of our UCSD Telescope data processing and visualization, we have dedicated use of several nodes on the SDSC Gordon supercomputer platform that stores and processes the indexed time-series data.
- We make use of an Energy Research Computing Allocations Process (ERCAP) Allocation at the National Energy Research Scientific Computing Center (NERSC) facility, a division of the Lawrence Berkeley National Laboratory located in Berkeley, California. SDSC has high bandwidth connectivity (10 GB) with the NERSC.GOV domain allowing regular file transfers for archival of historical data.

2. *System Inventory*

- Data Server
 - OS: FreeBSD 8.2
 - CPUs: 1 x 4 core Intel(R) Xeon(R) CPU E5620 @ 2.40 GHz
 - Memory: 6 GB RAM
 - Storage: 48 TB raw disk (38 TB after RAID 6 and ZFS overhead)

- Description: This machine is used as a second copy for scientific data replacing cloud.sdsc.edu.
- Web Server
 - OS: FreeBSD 9.2 (jail server)
 - CPUs: 2 x 6 core Intel(R) Xeon(R) CPU X5670 @ 2.93 GHz
 - Software: Apache
 - Memory: 48 GB RAM
 - Storage: minimum requirements for jailed operating system and mounted file systems.
 - Description: This server hosts FreeBSD jail services (that mount imported file systems from irori.caida.org) in support of the web infrastructure that serves the dataset description pages, forms, and project web pages <http://www.caida.org/data/> and <http://www.caida.org/projects/impact/>. We run two redundant jail servers so that one can take over if the other fails.
- Web Server
 - OS: FreeBSD 8.2 (jail server)
 - CPUs: 2 x 6 core Intel(R) Xeon(R) CPU X5670 @ 2.93 GHz
 - Software: Apache
 - Memory: 48 GB RAM
 - Storage: minimum requirements for jailed operating system and mounted file systems
 - Description: This machine is the primary jail server in support of the data distribution services.
- Data and Compute Server
 - OS: Ubuntu 14.04
 - CPUs: 2 x 4 core Intel(R) Xeon(R) CPU X5667 @ 3.07 GHz
 - Memory: 32 GB RAM
 - Storage: 12 TB of raw disk (~9 TB useable after RAID 5 and file system overhead)
 - Description: This host is the main data server for the Archipelago measurement infrastructure.
- Data Server
 - OS: FreeBSD 10
 - CPUs: 2 x 8 core ES-2650V2 @ 2.6 GHz
 - Memory: 256 GB RAM
 - Storage: Two disk trays: 94.5 TB (135 TB raw) & 126.4 TB (180 TB raw)
 - Description: This host acts as the primary data server and as management machine for the UCSD Network Telescope data.
- Data and Compute Server
 - OS: FreeBSD 10
 - CPUs: 1 x 6 core Intel E5-1650 3.5 GHz
 - Memory: 32 GB RAM
 - Storage: Two disk trays: 43 TB after RAID z3 and 2 hot spares (72 TB raw disk) & 57TB after RAID z3 (96TB raw).
 - Description: This server acts our primary data server.
- Data and Compute Servers
 - OS: Ubuntu Linux

- CPUs: 4x AMD Opteron(tm) Processor 6136 (quad-core 2.4 GHz)
- Memory: 64 GB RAM
- Storage: JBOD diskshelf: 48TB of raw disk with 37TB usable after raid 6 (planned for 2016)
- Description: We repurposed these two decommissioned SDSC Trestles nodes that now provide compute and storage resources in support of our topology experiments as well as our prototypical topology data query services.
- SDSC Gordon Supercomputer
 - OS: Rocks/CentOS
 - Dedicated Compute Nodes/CPU: 15 x Intel XEON E5 (Sandy Bridge) 2.6 GHz dual socket; 16 cores/node
 - Memory: 64 GB 1333 MHz RAM
 - Storage: 80 GB Intel SSD per node
 - Dedicated Flash-based I/O Nodes/CPU: 1 x 64 Intel Westmere; dual socket; 12 core
 - Memory: 4.8 TB Intel 710
 - Storage: SSD/node (300 TB total)
- NERSC HPSS Tape Archive Allocation
 - OS: AIX
 - Software: HPSS Tape Services
 - Description: For the period August 2014 – July 2015, we archived 133 TB of our historical UCSD Network Telescope (darknet) data using HPSS tape resources at the NERSC facility