

Supporting Research and Development of Security Technologies through Network and Security Data Collection

Cooperative Agreement FA8750-12-2-0326, CAIDA, UCSD

HOSTING INFRASTRUCTURE DESCRIPTION

The following describes the data hosting infrastructure deployed and maintained by CAIDA staff on the UC San Diego campus that supports datasets provisioned by CAIDA for the PREDICT Project. The described infrastructure was not exclusively funded by the DHS PREDICT Project.

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 (<http://cloud.sdsc.edu/>) 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

1. Data Server (thoth.caida.org)

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.

2. Web Server (alcatraz.caida.org)

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 from the thoth.caida.org data server to serve <http://data.caida.org/>, <http://topo-data.caida.org/>, and <http://imdc.datcat.org/>.

We run two redundant jail servers so that one can take over if the other fails.

3. Web Server (attica.caida.org)
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.
4. Web Server (indy.caida.org)
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.
5. Data Server (thor.caida.org)
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.
6. Data and Compute Server (irori.caida.org)
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.
7. Web Server (cider.caida.org)
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 <http://www.caida.org/data/> and <http://www.caida.org/projects/predict/>. We plan to switch to a jail server system next year.
8. Data Archival and Storage (cloud.sdsc.edu)
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.

9. 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.

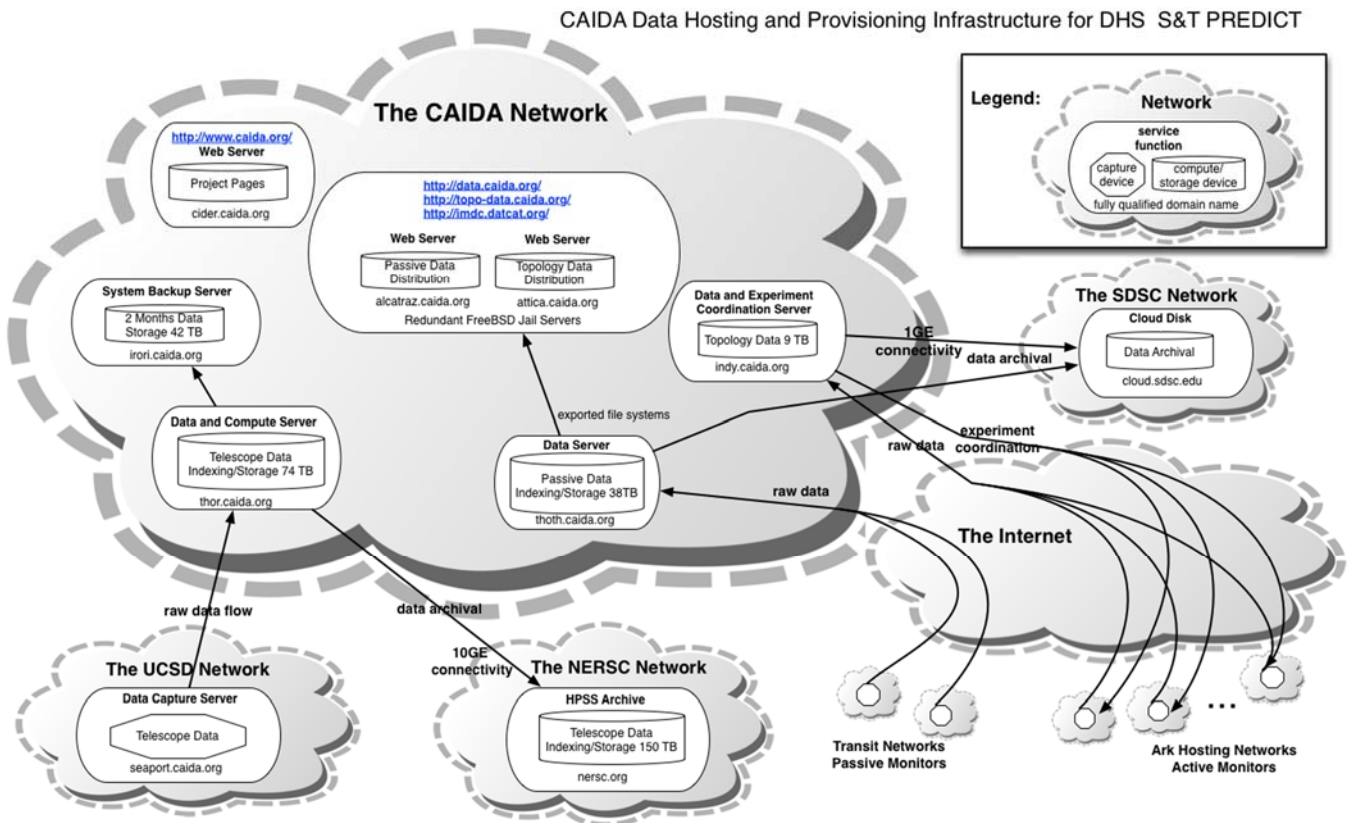


Figure 1. CAIDA UCSD Data Hosting and Provisioning Infrastructure for PREDICT.