Advancing Scientific Study of Internet Security and Topological Stability (ASSISTS)

Cooperative Agreement FA8750-18-2-0049, CAIDA, UCSD

TTA1 Capability Design Plan: HOSTING INFRASTRUCTURE DESCRIPTION

The following describes the data hosting infrastructure deployed and maintained by CAIDA staff on the UC San Diego campus in support of datasets provisioned by CAIDA for the DHS S&T IMPACT Project as of February 2018. The described infrastructure was not exclusively funded by the DHS IMPACT Project.

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 archiving, 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 (several hundred TB) locally attached disk systems that make use of the ZFS and Openstack Swift 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 well as servers running Ubuntu Linux.

We continue to use two repurposed nodes from the decommissioned SDSC Trestles Supercomputer 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 15 compute nodes and one I/O node on the SDSC Gordon supercomputer platform that stores and processes the indexed time-series data.

To store the growing volume of the UCSD Telescope 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.

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 used as a second copy for scientific data replacing cloud.sdsc.edu.

2. Web Server (charlotte.caida.org)
    OS: FreeBSD 10.3 (jail server)
    CPUs: 1 x 8 core Intel(R) Xeon(R) CPU E5-1660 @ 3 GHz
    Software: Apache

Memory: 32 GB RAM
Storage: 8 TB of RAID 6 for data and logs.
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/.

3. Web Server (attica.caida.org)
OS: FreeBSD 9.3 (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 https://data.caida.org/ and https://topo-data.caida.org/.

4. Data and Compute 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 (loki.caida.org)
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.

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: Two disk trays: 43 TB after RAID z3 and 2 hot spares (72 TB raw disk) & 114.5 TB after RAID z3 +1 hot spare (180 TB raw).
Description: This server acts our primary data server.

7. Data and Compute Servers (jones|junior.caida.org)
OS: Ubuntu Linux
CPUs: 4x AMD Opteron(tm) Processor 6136 (quad-core 2.4 GHz)
Memory: 64 GB RAM
Storage: JBOD disk shelf: 48TB of raw disk with 37TB usable after raid 6
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.

8. SDSC Gordon Supercomputer
OS: Rocks/CentOS

Dedicated Compute Nodes/CPUs: 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/CPUs: 1 x 64 Intel Westmere; dual socket; 12 cores
Memory: 4.8 TB Intel 710
Storage: SSD/node (300 TB total)

9. NERSC HPSS Tape Archive Allocation
OS: AIX
Software: HPSS Tape Services
Description: For the period August 2015 – July 2016, we archived 274 TB of our historical UCSD Network Telescope (darknet) data using HPSS tape resources at the NERSC facility.

10. Object Storage Server (hermes.caida.org)
OS: Ubuntu 16.04
CPUs: 1 x 16 core Intel ® Gold 6142 @ 2.6 GHz
Memory: 128 GB RAM
Storage: Two disk trays: one with 24 x 8 TB drives, one with 45 x 10 TB drives.
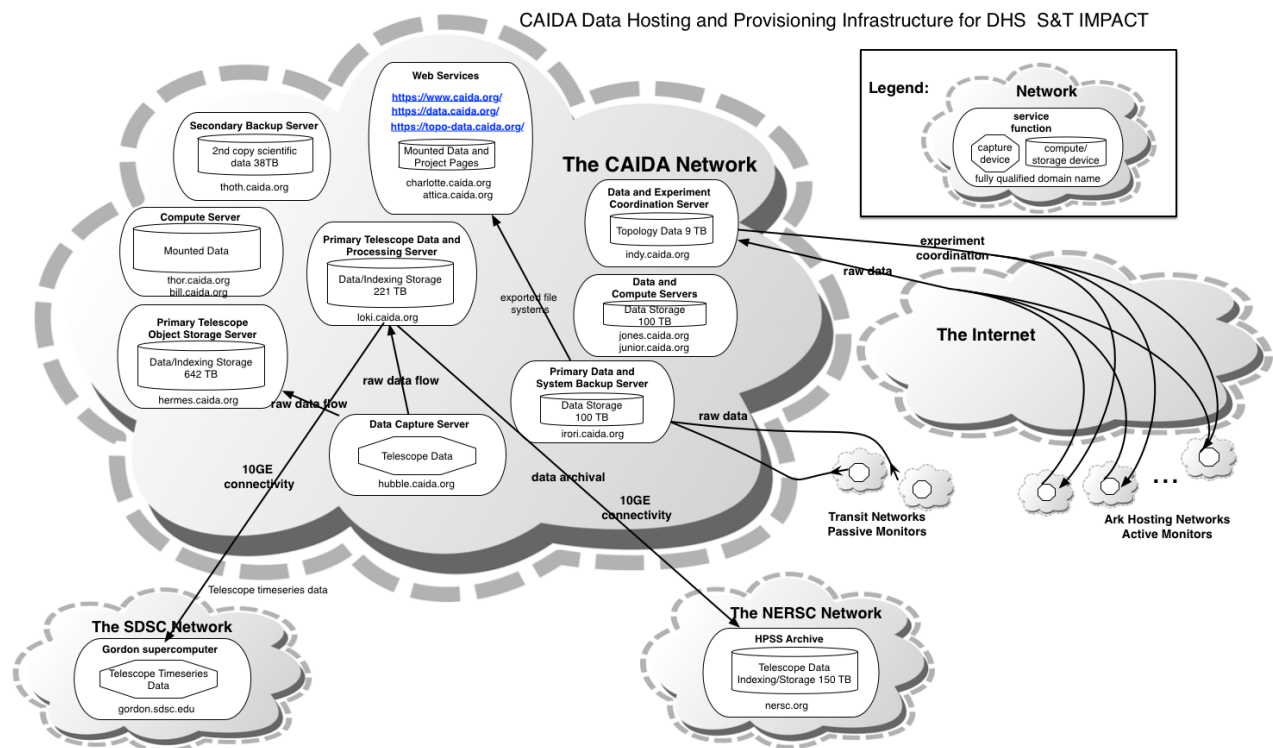Description: This acts as the swift object storage server for telescope infrastructure.

Figure 1. CAIDA UCSD Data Hosting and Provisioning Infrastructure in support of DHS S&T IMPACT Project (February 2018).