Hawaii Pan-STARRS Data Movement Issues Summary

Contact Point: Hans Addleman (addlema@iu.edu)
Audience: General
Last updated: August 17, 2022

The Panoramic Survey Telescope and Rapid Response System (Pan-STARRS) program shares approximately 100 terabytes of data yearly between the Institute for Astronomy (IfA) at the University of Hawaii (UH) and the Space Telescope Science Institute (STSCI) at John Hopkins University in Baltimore, Maryland. This enables researchers to more accurately estimate galaxy redshifts, improving their understanding of the local cosmic expansion and dark energy. In November 2018, engineers from UH approached engineers at Indiana University (IU) for assistance because they had experienced a diminished maximum transfer rate of only 320 Mbps, a fraction of the estimated full capacity between IfA and STSCI, which was believed to be equipped with 10Gbps and 100Gbps networks. Aiming to achieve transfer rates in the multi gigabit range, UH sought IU's assistance in debugging the path and recommending a parallel file transfer tool.

Over the next 3 months, the International Research Network Connections (IRNC) NOC, the IRNC Performance Engagement Team (PET) and International Networks at Indiana University (IN@IU) staff worked closely with network engineers and IT staff from University of Hawaii, John Hopkins University, Indiana University, Internet2, and the Mid-Atlantic Crossroads (MAX) Gigapop (which supports R&E networking in the Maryland, Virginia, and DC area) to actively troubleshoot the issue, identify bottlenecks, and resolve the identified problems.

The team made heavy use of perfSONAR nodes for ongoing and adhoc testing during the engagement across the full path. These nodes were located at various points of both end networks and the Wide Area Network (WAN) between them, including Internet2, TransPAC, and MAX gigapop nodes. UH moved multiple perfSONAR nodes around their network to improve the accuracy of testing.I.

The engagement identified a number of issues. Some of these issues were solved during testing and others were cataloged for addressing at a later date. On the University of Hawaii side, these included:

- The Top of the Rack (TOR) switches in the UoH data center Science DMZ were determined to be underpowered for the level of data transmission they were experiencing, so the critical data servers were moved.
- Misconfigured access control lists and firewalls in the Science DMZ also contributed to the poor performance. IfA worked to eliminate these bottlenecks by redesigning the equipment layout so that data transfer nodes were not behind the firewall.
- The default routing between the UH hosts and the JHU hosts showed a suboptimal, longer route through the Internet2 Network. UH staff moved peering to their PIREN 100G link to Los Angeles and this allowed the traffic to take an optimal path from end to end.

This material is based upon work supported by the National Science Foundation under Grant No. 1826994
info@epoc.global

On the JHU side, these included:

- The determination that the JHU portion of the path was actually only a 1G path between JHU and the MAX gigapop. The internal network was upgraded to 10G from the end data receiver host through to the Internet2 connection via the MAX Gigapop.
- perfSONAR nodes were installed inside and outside the JHU firewall to enable better ongoing testing and error identification..

Both UH and JHU found the following issues with their data transfer methodology and hardware:

- Across the end-to-end path, the Maximum Transmission Unit settings on all of the routers and transfer hosts were upgraded to 9000 byte size frames (Jumbo Frames). This improved network performance by making data transmissions more efficient, because the CPUs on switches and routers can process a larger payload for each frame. This method only works if each link in the network path, as well as the servers and endpoints, is configured to enable jumbo frames at the same MTU size.
- On both ends, the TCP Buffer settings for the end hosts were misconfigured for large scale data transfers, so these also were updated to the ESnet recommended settings (http://fasterdata.es.net/host-tuning/background/).

In addition, due to the age of the software and system set up for this collaboration, several inefficiencies were identified:

- Because the software was bespoke to PanSTARRS, and written over many years, some aspects of it were ineffective for today's systems. Specifically, the system required manual intervention at various points which could delay the workflow. Full resolution will require significant re-write of the workflow tool.
- Within the bespoke software framework, file transfers were delayed by a per file DNS lookup that would hang due to a misconfiguration of the Web Proxy piece of the file transfer mechanism. The configuration of the proxy was updated to resolve this issue.
- The collected data did not exist in a single location. Instead, it was spread out across over 160 discrete logical storage volumes on 32 hosts, many of which had not been tuned (or could not be tuned) to enable fast data transfers. In addition, some of the hosts had aged to the point of being unreliable and could crash in the middle of data access actions. The project is working towards a new, unified storage system on modern equipment to address these issues.

The file transfer was fully re-evaluated in June of 2018 and after our engagement saw a 3x jump in overall performance, reporting a sustained 1Gbps transfer rate, up from the original 320Mbps.

The results of this engagement also led the PIREN project to receive a supplemental National Science Foundation (NSF) award to enhance the capabilities of their overall network, Science DMZ, Data Transfer hardware, and network testing hardware. They are currently working to procure, design, and put this new architecture in place. IN@IU and the PET will remain engaged as needed.

This material is based upon work supported by the National Science Foundation under Grant No. 1826994
info@epoc.global