EPOC Advice on 100G perfSONAR Nodes

Contact Point: Jason Zurawski (zurawski@es.net)

Audience: General

]Last updated: August 17, 2022

ABOUT EPOC

Over the last decade, the scientific community has experienced an unprecedented shift in the way research is performed and how discoveries are made. Highly sophisticated experimental instruments are creating massive datasets for diverse scientific communities and hold the potential for new insights that will have long-lasting impacts on society. However, scientists cannot make effective use of this data if they are unable to move, store, and analyze it. The Engagement and Performance Operations Center was established in 2018 as a collaborative focal point for operational expertise and analysis and is jointly led by Indiana University (IU) and the Energy Sciences Network (ESnet). EPOC provides researchers with a holistic set of tools and services needed to debug performance issues and enable reliable and robust data transfers. By considering the full end-to-end data movement pipeline, EPOC is uniquely able to support collaborative science, allowing researchers to make the most effective use of shared data, computing, and storage resources to accelerate the discovery process.

EPOC supports six main activities:

- **Roadside Assistance and Consultations** via a coordinated Operations Center to resolve network performance problems with end-to-end data transfers;
- **Application Deep Dives** to work more closely with application communities and understand full workflows for diverse research teams in order to evaluate bottlenecks and potential capacity issues;
- **Network Analysis enabled by the NetSage** monitoring suite to proactively discover and resolve performance issues;
- **Data Transfer Testing/ Data Mobility Exhibition** to check transfer times against known good end points;
- **Provision of managed services** via support through the IU GlobalNOC and our Network Partners;
- **Coordinated Training** to ensure effective use of network tools and science support.

The Request

As part of the EPOC consultation process, we are often contacted for advice in using or deploying 100G perfSONAR nodes. Our response generally begins by suggesting that the folks take a step back and first identify their use cases versus making a choice based on technology availability, and to take a closer look at the networks they're trying to test. One mis-step that sites make when looking at 100G perfSONAR nodes is being tempted to build the biggest they can, with the belief that bigger is always better. However, our experience has shown this approach is often counterproductive for several reasons that are apparent when you examine the environment in more detail.

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

Why 100G Isn't the Right Solution

Regular 100G testing is still not a very common thing. It is true people have capability, but two things happen when a 100G node is available:

People use the perfSONAR lookup service and test to it (because they can), which can generate 100G traffic bursts at a campus, through a regional/through a backbone, or across an exchange point.

People may not know a resource is100G capable and test against it from their side with a smaller tester (100M/1G/10G/40G). When this happens, the results include microburst loss events, errors, and measured results that are unclear and not indicative of true performance

A perfSONAR node that can run a 100G test will also be significantly more expensive than one that only runs 1 or 10G tests,

Our Current Recommendations

We advocate, as does the perfSONAR consortium (http://perfsonar.net), to build more smaller (e.g. 10G) perfSONAR nodes that can be deployed in multiple places as opposed to one giant node that is only in one place. Often having more testers available is far more valuable in practice.

In terms of the hardware, we recommend construction that can scale with the use case:

CPUs that are high clock rate, so . 3.6Ghz or above, with a moderate core count. A fast CPU is critical to TCP working effectively, and multiple cores will ensure that one core is pinned to testing, which is still a serial operation, and then the others can handle needed system operation (I/O, graphical displays like maddash, etc).

Maximize the use of the RAM slots when applicable. This doesn't mean getting the largest capacity for each DIMM, but it does mean filling all available RAM sockets with equal size DIMMs. The equal size is important as it greatly helps the machine to balance memory load.

Main storage can be conservative. For perfSONAR nodes, regular spinning disks for the OS and measurement storage are fine. Alternatively, it is possible to configure a central collector and not worry about local storage at all. Unlike when building a Data Transfer Node (DTN), storage doesn't need to be NVMe/VROC or any more performant variety.

Use a network card (or cards) that facilitate swappable optics and multiple Ethernet standards. For example, several available options can facilitate 10/40/50/100G operation, depending on the optic or cable used in the QSFP28 port. This way it is possible to have 100G capability if absolutely needed on occasion, but it can be left at a lower more friendly speed at other times.

The advantage of having multiple smaller perfSONAR test points at different locations in the network cannot be overstated.

For more information on how ESnet and the perfSONAR consortium design the perfSOSNAR test points, we recommend:

perfSONAR Hardware Requirements: https://docs.perfsonar.net/install_hardware.html

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

info@epoc.global

General information about perfSONAR: https://www.perfsonar.net/

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