

EPOC Application Deep Dives

Last updated Feb 13, 2019

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 five main activities

- *Roadside Assistance* via a coordinated Operations Center to resolve network performance problems with end-to-end data transfers reactively;
- *Application Deep Dives* to work more closely with application communities to 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;
- Provision of *managed services* via support through the IU GlobalINOC and our Regional Network Partners;
- *Coordinated Training* to ensure effective use of network tools and science support.

APPLICATION DEEP DIVES

Whereas the Roadside Assistance portion of EPOC can be likened to calling someone for help when your car breaks down, EPOC has also implemented a process similar to the yearly tuneup, which we refer to as *Deep Dives*. Deep Dives aim to understand the full science pipeline for research teams and suggest alternative approaches for the scientists, local IT support, and national networking partners as relevant to achieve the long term research goals via workflow analysis, storage/computational tuning, identification of network bottlenecks, etc.

The Deep Dive approach is based on an almost 10-year practice used by ESnet to understand the growth requirements of DOE facilities

(<https://fasterdata.es.net/science-dmz/science-and-network-requirements-review/>). We have adapted this approach for work with individual science groups. The discussion, which takes

place through a set of structured data-centric conversations and questionnaires, includes:

- Background information, including a brief overview of the science being performed and where it sits within its discipline.
- Identification of collaborators - both people and institutions that a science group interacts with.
- How the team is currently using both local and remote scientific instruments and facilities.
- The process of the science, which includes explicitly explaining 'a day in the life' of the science group. This will tie together the the people/collaborators, the use of resources, and the science goals for the team.
- A description of the software infrastructure that is supporting the research goals.
- Details about the network and data architecture between the collaboration sites, as part of the science workflow.
- The use, if any, of cloud services.
- A discussion of any outstanding issues and pain points.

At the end of the interaction with the research team, EPOC staff can come away with a good understanding of the research, data movement, who's using what pieces, dependencies, and time frames. This enables us to identify possible bottlenecks or areas that may not scale in the coming years, and to point the research team towards existing resources so that they can reach their goals more effectively.