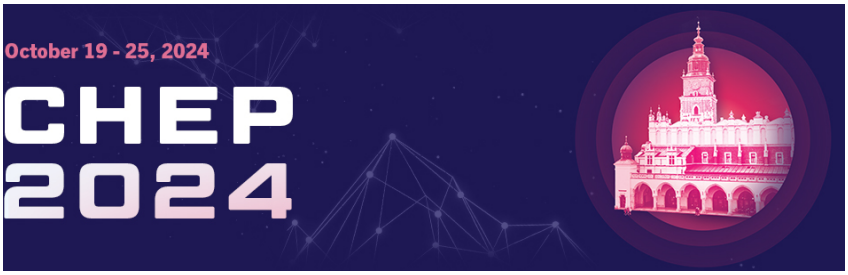


Contribution ID: **227** Type: **Talk**

# Towards an Introspective Dynamic Model of Globally Distributed Computing Infrastructures

*Thursday 24 October 2024 14:24 (18 minutes)*

Large-scale scientific collaborations like ATLAS, Belle II, CMS, DUNE, and others involve hundreds of research institutes and thousands of researchers spread across the globe. These experiments generate petabytes of data, with volumes soon expected to reach exabytes. Consequently, there is a growing need for computation, including structured data processing from raw data to consumer-ready derived data, extensive Monte Carlo simulation campaigns, and a wide range of end-user analysis. To manage these computational and storage demands, centralized workflow and data management systems are implemented. However, decisions regarding data placement and payload allocation are often made disjointly and via heuristic means. A significant obstacle in adopting more effective heuristic or AI-driven solutions is the absence of a quick and reliable introspective dynamic model to evaluate and refine alternative approaches.

In this study, we aim to develop such an interactive system using real-world data. By examining job execution records from the PanDA workflow management system, we have pinpointed key performance indicators such as queuing time, error rate, and the extent of remote data access. The dataset includes six months of activity. Additionally, we are creating a generative AI model to simulate time series of payloads, which incorporate visible features like category, event count, and submitting group, as well as hidden features like the total computational load—derived from existing PanDA records and computing site capabilities. These hidden features, which are not visible to job allocators, whether heuristic or AI-driven, influence factors such as queuing times and data movement.

**Primary authors:** REN, Yihui (Brookhaven National Laboratory (US)); Dr KILIC, Ozgur (Brookhaven National Laboratory); Dr PARK, David (Brookhaven National Laboratory); KORCHUGANOVA, Tatiana (University of Pittsburgh (US)); Dr SUTER, Frederic (Oak Ridge National Laboratory); BOUDREAU, Joseph (University of Pittsburgh (US)); PODHORSZKI, Norbert (Oak Ridge National Laboratory); NILSSON, Paul (Brookhaven National Laboratory (US)); Dr SRI VATSAVAI, Sairam (Brookhaven National Laboratory); KLASKY, Scott; CHOWDHURY, Tasnuva (University of the Witwatersrand (ZA)); Mr FENG, Shengyu (Carnegie Mellon University); MARTINEZ OUTSCHOORN, Verena Ingrid (University of Massachusetts (US)); Prof. YANG, Yiming (Carnegie Mellon University); MAENO, Tadashi (Brookhaven National Laboratory (US)); KLIMENTOV, Alexei (Brookhaven National Laboratory (US)); Dr HOISIE, Adolfy (Brookhaven National Laboratory)

**Presenter:** Dr PARK, David (Brookhaven National Laboratory)

**Session Classification:** Parallel (Track 7)

**Track Classification:** Track 7 - Computing Infrastructure