BNL’s future computational fabric for HL-LHC and EIC

BNL as host to cutting edge user facilities as well as US participation in large scale international science has built scientific computing as a lab core capability. In nuclear and particle physics (NPP), the RHIC and ATLAS programs have made BNL a world leader in data intensive scientific computing. The decision to site the Electron Ion Collider (EIC) at BNL will make the next decades just as transformative. The EIC from the 2030s will enable precision measurements of the 3D structure binding the constituents of protons and nuclei, systematically exploring the QCD bound state. The EIC's computing needs will draw on a computing fabric progressively enhanced by the coming NPP program. RHIC's sPHENIX experiment in 2023-2025 will make pioneering use of high-rate, high-volume data streaming, producing data volumes approaching the HL-LHC, which starts operating in 2027.

This 30+ year NPP science program will transform BNL’s computing fabric in ways that will both grow the lab’s leading role in scientific computing and drive new capabilities applicable beyond NPP, as data-driven techniques continue to grow exponentially across the sciences and society. Key aspects of the science which will strongly impact computing include

- **Data intensity.** The HL-LHC and EIC rely on high-luminosity, high-throughput data streams to reach the sensitivity required by small signal-to-background and a precision science program. Exabyte data volumes must be accessed by scientists around the world in a prompt and cost-effective way. Economizing costly storage and utilizing powerful networks for efficient remote data access are today and will remain primary challenges in building the computing fabric our data intensive science requires.
- **Reconstruction.** Detector design is reaping the benefits of modern processing power. Detector granularity is growing as is the complexity of the data. Compute demands are large; high concurrency and machine learning (ML) will be central in meeting them. Traditional methods will increasingly fail to perform well on new architectures, contributing to ever increasing use of AI/ML. This rapidly evolving environment will have to be tracked and supported by the compute fabric and expertise base.
- **Simulation.** Simulation will be a primary compute component for the HL-LHC and EIC. ML will be a powerful means of speeding the simulation on modern architectures without compromising physics quality, with ML models based on large samples of detailed full simulation and real data. This is another driver for the compute fabric effectively becoming a ML engine.
- **Data analysis.** The EIC and HL-LHC era analyst will work within a distributed computing fabric that abstracts away the physical resources, on analyses defined and developed to be automatically reproducible, preservable and re-interpretable. Data processing, analysis and theory will be closely coupled, with close connections between analysis and theoretically derived quantities like PDFs. The publishing of data, simulations, likelihood analyses, ML models will serve enduring collaborations between experimentalists and theorists. As new theoretical interpretations arise and new analysis methods emerge from domains such as AI/ML and ultimately quantum, it must be possible to apply them to existing experimental datasets. A priority for the fabric should be to serve the full gamut from data acquisition to preserved analysis to theoretical (re)interpretation, persistently over time.