



Contribution ID: 264 Contribution code: TUE 22

Type: Poster

Investigating Data Access Models for ATLAS: A Case Study with FABRIC Across Borders and ServiceX

Tuesday 22 October 2024 16:00 (15 minutes)

This study explores possible enhancements in analysis speed, WAN bandwidth efficiency, and data storage management through an innovative data access strategy. The proposed model introduces specialized “delivery” services for data preprocessing, which include filtering and reformatting tasks executed on dedicated hardware located alongside the data repositories at the CERN Tier-0 or at Tier-1 or Tier-2 facilities. Positioned near the source storage, these services are crucial for limiting redundant data transfers and focus on sending only vital data to distant analysis sites, aiming to optimize network and storage use at those sites. Within the scope of the NSF-funded FABRIC Across Borders (FAB) initiative, we assess this model using an “in-network, edge” computing cluster at CERN, outfitted with substantial processing capabilities (CPU, GPU, and advanced network interfaces). This edge computing cluster features dedicated network peering arrangements that link CERN Tier-0, the FABRIC experimental network, and an analysis center at the University of Chicago, creating a solid foundation for our research.

Central to our infrastructure is ServiceX, an R&D software project under the Data Organization, Management, and Access (DOMA) group of the Institute for Research and Innovation in Software for High Energy Physics (IRIS-HEP). ServiceX is a scalable filtering and reformatting service, designed to operate within a Kubernetes environment and deliver output to an S3 object store at an analysis facility. Our study assesses the impact of server-side delivery services in augmenting the existing HEP computing model, particularly evaluating their possible integration within the broader WAN infrastructure. This model could empower Tier-1 and Tier-2 centers to become efficient data distribution nodes, enabling a more cost-effective way to disseminate data to analysis sites and object stores, thereby improving data access and efficiency. This research is experimental and serves as a demonstrator of the capabilities and improvements that such integrated computing models could offer in the HL-LHC era.

Primary authors: JORDAN, David (University of Chicago (US)); HU, Fengping (University of Chicago (US)); VUKOTIC, Ilija (University of Chicago (US)); STEPHEN, Judith Lorraine (University of Chicago (US)); BRYANT, Lincoln (University of Chicago (US)); GARDNER JR, Robert William (University of Chicago (US)); MC KEE, Shawn (University of Michigan (US))

Presenter: VUKOTIC, Ilija (University of Chicago (US))

Session Classification: Poster session

Track Classification: Track 1 - Data and Metadata Organization, Management and Access