Contribution ID: **744** Contribution code: **contribution ID 744**   Type: **Poster**

ServiceX: Making all Data Columnar

ServiceX is a cloud-native distributed application that transforms data into columnar formats in the python ecosystem and ROOT framework. Along with the transformation, is applies filtering, and thinning operations to reduce the data load sent to the client. ServiceX, designed for easy deployment to a Kubernetes cluster, is runs near the data, scanning TB's of data to send GB's to a client or analysis facility. In parallel it can quickly read data from a variety of formats, apply selection criteria, calculations, sorting operations. Adaptors are available for ROOT and parquet files, as well as awkward arrays and ROOT's RDataFrame interface. An overview of ServiceX, its connections inside and outside of Particle Physics, and the concepts behind transformation and applicability to data preservation will be described. Open data from ATLAS run 1 (simple ROOT TTree files) and CMS Run 1 AOD (complex binary datafiles) will be used as examples to demonstrate the functionality.

Significance

- First time able to service root files, parquet files, and as awkward arrays, or feed into RDataFrame
- Installations at various Analysis Facilities have now occurred
- Gained users from the Dark Matter Community (which will discuss briefly here)
- Can service modern Run 2 and very old Run 1 data to the same set of tools

References

Speaker time zone

Compatible with Europe

Authors: ECKART, Andrew (University of Chicago); GALEWSKY, Benjamin (Univ. Illinois at Urbana Champaign (US)); WATTS, Gordon (University of Washington (US)); NEUBAUER, Mark Stephen (Univ. Illinois at Urbana-Champaign); VUKOTIC, Ilija (University of Chicago (US)); CHOI, Kyungeon (University of Texas at Austin (US)); PROFFITT, Mason (University of Washington (US)); Mr DECHEINE, Nick (Wisconson); ONYISI, Peter (University of Texas at Austin (US)); GARDNER JR, Robert William (University of Chicago (US)); THAPA, Suchandra (University of Chicago); THAPA, Suchandra (University of Chicago)

Presenter: WATTS, Gordon (University of Washington (US))

Session Classification: Posters: Walnut

Track Classification: Track 2: Data Analysis - Algorithms and Tools