Contribution ID: **426** Contribution code: **WED 11**   Type: **Poster**

Comprehensive Description and Orchestration of Complex Data Processing Pipelines

Wednesday 23 October 2024 16:00 (15 minutes)

Managing and orchestrating complex data processing pipelines require advanced systems capable of handling diverse and collaborative components, such as data acquisition, streaming, aggregation, event identification, distribution, detector calibration, processing, analytics, and archiving. This paper introduces a data processing workflow description and orchestration system designed to facilitate the coordination and operation of these components using both centralized orchestration and decentralized choreography approaches. Our system employs a decentralized actor-based model used in the data acquisition system and data stream processing framework at Jefferson Lab (JLAB) to create component-specific configurations for the effective choreography of component actors. Simultaneously, the centralized orchestration provides global control and management of the entire data processing pipeline from acquisition to final processing. Our system's core is an ontology language developed explicitly for serializing data processing pipeline descriptions. A user-friendly graphical interface also enables seamless data pipeline composition and real-time monitoring. This integrated approach ensures efficient deployment, management, and orchestration of data processing workflows, ensuring robustness and flexibility in handling complex scientific data workflows.

Primary author: Dr GYURJYAN, Vardan

Presenter: Dr GYURJYAN, Vardan

Session Classification: Poster session

Track Classification: Track 4 - Distributed Computing