Wednesday, August 17, 2022
3:00 PM – 3:20 PM
Universal Data Distribution

Rick Taylor
Solutions Engineer
Cloudera Government Solutions, Inc.

Abstract:
The Army needs a reliable, scalable data transport mechanism to deliver sensor data from origination through all points of consumption; at the edge, on-premise, and in the cloud in a simple, secure, universal, and scalable way. The data transport architecture should provide guaranteed delivery, even in contested or denied, disrupted, intermittent or limited bandwidth (DDIL) environments. These requirements can be addressed by a Universal Data Distribution (UDD) architecture.

UDD provides a wide range of extensible capabilities ranging from real-time data ingestion, edge processing, transformation, and routing through to descriptive, prescriptive, and predictive analytics. UDD provides the capability to connect to any data source anywhere, with any structure, process it, and reliably deliver prioritized sensor data to any destination. Data can be securely shared across on-premises, public cloud, hybrid cloud environments and multiple enclaves. The movement of data across multiple enclaves is facilitated by utilization of any DoD approved Cross Domain Solution. Once the architecture is set up to ingest sensor data from thousands of endpoints into the UDD, the system can scale up or out based on mission requirements. In fact, one of our customers is ingesting data from over 100,000 endpoints.

The UDD can act as a neutral, bidirectional, data movement engine between sensors and any of the Army's data platforms. Sensor data from thousands of end points can provide valuable operational insights about threats and vulnerabilities resulting in improved battle space awareness. UDD is designed with ease of use in mind. Data transfer requirements are instantiated using a low code environment to specify and automate reliable data movement. UDD offers a flow-based, low-code development paradigm that provides the best impedance match with how developers design, develop, and test data distribution pipelines. With over 400 connectors and processors, UDD enables a broad range of data distribution capabilities. Mission specific, custom processors can also be developed as needed. These data distribution flows can be version controlled into a catalog where operators can self-serve deployments to different runtimes.

Cloudera UDD provides secure, universal, hybrid, and streaming data distribution enabling the Army to effectively manage data flows and sensor feeds. Cloudera's technology solutions enable organizations to capture, store, analyze and act on any data at massive speed and scale with our Cloudera Data Flow (CDF), Cloudera Data Platform (CDP), and Cloudera Machine Learning (CML) solutions. CDF is the enabling technology for the UDD architecture. Cloudera solutions are built upon proven open source components and open standards with wide industry adoption, they are readily implemented and straightforward to integrate with other platforms. These capabilities and solutions ultimately enable US Military and Civilian leaders and analysts with the real-time and predictive insights they need to meet their objectives, and support the mission. CDF, CDP and CML provide a comprehensive data ecosystem

for security and governance of mission data, as well as a rich toolset for its operational use. Cloudera technologies are accredited in the IC and DoD communities.