Thursday, August 21, 2025

12:00 PM - 12:20 PM

Real-Time Data for Mission Analytics

Carmelo McCutcheon

CTO

VAST Federal

Abstract:

This presentation addresses Problem Statement 3 posed by the Army Cyber Center of Excellence (CCoE): "How can technology ingest data from multiple sensors, geotag data, and aggregate for analysis? Can technology do predictive analysis based on preconditioned filters? Can this technology integrate data into mission command systems?"

Our solution demonstrates a comprehensive end-to-end data analytics pipeline specifically designed for mission-critical military applications, leveraging the VAST Data Platform as the foundational infrastructure integrated with Apache Spark, Trino query engine, and Apache Superset. This architecture directly addresses the Army's need for real-time sensor data processing, geospatial intelligence, and seamless integration with existing mission command frameworks.

Technical Architecture Overview

The implemented solution features a four-stage pipeline architecture that begins with direct sensor data capture through high-throughput ingestion mechanisms capable of processing multiple sensor formats simultaneously. Raw sensor streams flow through automated triggers and database functions for initial data quality assurance and normalization, ensuring consistent data structures before entering the transformation layer.

Apache Spark serves as the primary ETL engine, performing complex data transformations, feature engineering, and real-time geospatial tagging operations on streaming sensor inputs. The VAST DataBase provides unified storage, database, and stream processing capabilities in a single all-flash system optimized specifically for AI and analytics workloads required in tactical environments.

Predictive Analytics and Mission Integration

The system integrates Trino's distributed SQL query engine running natively on VAST's serverless compute infrastructure, enabling high-performance interactive analytics on massive datasets without data movement penalties. This integration supports both real-time and historical data queries, facilitating predictive analytics through machine learning models that adapt to evolving operational conditions.

Preconditioned filtering capabilities are achieved through configurable data preprocessing pipelines that implement automated anomaly detection, data normalization, and feature selection processes. Mission command system integration is facilitated through RESTful APIs and standardized military data formats, enabling seamless incorporation of analytical outputs into existing C2 frameworks used by the Army Cyber Center of Excellence2.

Visualization and Decision Support

Apache Superset interfaces with Trino and the VAST DataBase to provide interactive dashboards and mission-critical insights to command personnel. The visualization layer supports real-time monitoring of sensor networks, predictive threat assessments, and geospatially-aware tactical displays that enhance situational awareness for commanders operating in contested environments34.

Performance evaluations demonstrate the system's ability to process high-velocity sensor streams with sub-second latency while maintaining data accuracy and supporting concurrent analytical workloads, making it well-suited for deployment in mission-critical environments requiring continuous monitoring and intelligent decision-making capabilities.

This unified architecture eliminates traditional data silos and reduces operational complexity by providing a single platform for storage, processing, and analysis, directly supporting the Army's multi-domain operations requirements and cyber electromagnetic activities as emphasized at TechNet Augusta.