GSAW 2024 Tutorial E: Half Day

Modern Data Engineering for a Data-Centric Ground Enterprise

Overview:
A data centric architecture is defined as an architectural approach where data is separated from their applications and securely made data available to a broad range of tools and analytics within and across domains for enrichment and discovery. [Source: IC Data Management Lexicon]

In an application centric enterprise, data is often locked in their respective mission applications and not able to be shared due to a myriad of interoperability and integration challenges. In recent years, the problem is further exacerbated by the exponentially growing volume, velocity, and variety of data. Government organizations are coming to the realization that enterprise architecture, system acquisition, and mission operations need to shift from an application centric mindset to a data centric mindset to free the data for enterprise access to authorized consumers for multiple uses and re-use, resulting in a need to follow a set of data-centric principles [Source: The Data Centric Manifesto], such as:

- Data is a key asset of any organization, even as applications come and go
- Data is self-describing and does not rely on an application for interpretation and meaning.
- Data is expressed in open, non-proprietary formats.
- Access to and security of the data is a responsibility of the data layer, and not managed by applications.

This tutorial lays out a conceptual framework for architecting, engineering, and operating a data centric enterprise. Much of the framework falls under an increasingly important discipline of data engineering. Data engineering is primarily concerned with the theory, methods, algorithms and models that extract knowledge and insights from data.

This tutorial will touch on the data engineering topics outlined below:

1. Introduction and the Case for Data Centricity
2. Data Pipeline Concepts for Processing and Transformation
3. Data Store and Data Exposure Services Concepts
4. Dataset Registration, Cataloging, and Discovery Concepts
5. Data Tagging and Metadata Management Concepts
6. Data Understandability and Interoperability Concepts
7. Data Access and Usage Concepts
8. Digital Policy Concepts for Data Access
9. Data Analysis and Exploitation Concepts
10. DataOps Concepts
11. Data Measurement and Tracking Concepts
12. Governance of Data Lifecycle Operations Concepts

Instructors:
Eric Yuan and Victor Rohr, The Aerospace Corporation

Biographies
Dr. Eric Yuan has over 20 years of industry experience and has been with Aerospace since 2012. Prior to working with Aerospace, Dr. Yuan was previously with Booz Allen Hamilton, Oracle, Broadband Office, and Bell Atlantic. Dr. Yuan has a B.S in Computer Science, M.S. in Systems Engineering from University of Virginia and a Ph.D. in Computer Science from George Mason University.

Mr. Victor Rohr is a Senior Project Engineer with the Aerospace Corporation, having joined the Aerospace team in 2003 after a 21 year career in the United States Navy. During the last 12 years of his Navy career, Mr. Rohr served as the Information Systems Director for a US Navy Command and was responsible for database architecting, software development and numerous technical and programmatic responsibilities. Within the Aerospace Corporation, he has held several positions spanning highly technical and programmatic roles, leading to his recognition as a leader in the drive for IC-wide semantic consistency.

Description of Intended Audience and Recommended Prerequisites:
No pre-requisites required.

What can Attendees Expect to Learn:
- Key concepts for a data-centric architecture (DCA)
- Data engineering principles across the entire data lifecycle
- Key building blocks of a data-centric ground enterprise, from both operational and technical perspectives
- Implementation considerations that may help guide technical solutions