denodo

How to Accelerate AI with a Logical Data Fabric

Data management and data governance are foundational to AI. You can have the greatest algorithms in the world and the best Large Language Models (LLMs), but without proper data to train them or add context, the effort will fall short.

At a high level, government officials recognize this. In a March 2024 memorandum titled "Advancing Governance, Innovation, and Risk Management for Agency Use of Artificial Intelligence," the Office of Management and Budget calls for agencies to ensure that "the data used to develop and test AI are appropriately inventoried, shared, and released" in agency repositories. And as part of its Data, Analytics and Artificial Intelligence Adoption Strategy, the Department of Defense put forward its VAULTIS framework, calling for data to be visible, accessible, understandable, linked, trustworthy, interoperable and secure.

Likewise, the General Services Administration's AI Guide for Government points to data collection, ingestion, management and manipulation as some of an AI implementation's critical aspects. But agencies now need to go beyond this conventional approach and focus on the entirety of data lifecycle management.

AI initiatives often need to draw from multiple data sources. For example, workforce data seems like it would come from one source, but it is actually scattered across multiple databases, making it a challenge to use AI to improve human resources and talent management functions. Data science teams can do this, but conventional approaches make it labor-intensive and time-consuming.

Ultimately, business leaders across government need to get answers quickly to their most pressing problems. AI can help, but to use it, agencies first must improve existing data access strategies.

A platform that delivers a logical data fabric can help accelerate AI across the federal landscape.

The Challenge: A Fragmented Data Landscape

As government looks to take advantage of AI's game-changing promise, several challenges stand out.

A distributed environment: Data lives in many places within and across agencies and reaching across those varied repositories to access the information needed to inform AI models can be tough. In sharing that distributed data, agencies may need to replicate it, and every time it is copied, maintaining a single source of truth gets harder. This complicates data management, governance and security, and heightens the risk of noncompliance with regulations, as tracking and controlling sensitive information becomes tricky. All this adds cost and complexity and leads agencies to closely guard the ownership of their data, which complicates their AI efforts.

Siloed systems: Data is not only distributed but siloed, kept in distinct repositories that often cannot communicate. What's more, the data often exists in different formats, so even if it's accessible, it's inconsistent. Whether in legacy or modern systems, data in silos isn't readily usable by end users — in this case, the AI models or LLMs that needed to be informed by that data.

Ridged integration approaches: Agencies widely and effectively use data warehouses, data lakes and similar storage solutions. However, they rely on rigid extract, load and transform processes or scripts, which can be time-consuming to develop or modify. Because these scripts process data in batches, they also introduce latency. These traditional approaches have their place, but they fall short when it comes to meeting the fast-evolving demand for timely data to be used by AI. In today's landscape, relying solely on these methods for data integration and management just isn't enough.

The Solution: A New Approach to Data Management and Delivery

The Denodo Platform from Denodo, a leading data management company, brings a new approach to solving these problems and accelerating government's AI efforts. It features:

A Logical Data Fabric

A logical data fabric enables enterprises to integrate, manage and deliver their distributed data to any user in real time. Even in a distributed data landscape, this approach makes it easy to access, monitor and manage data, with increased access to new data sources. Automation and augmentation support real-time access, enabling data engineering teams to put the data to work more quickly and effectively.

Advanced semantics simplify metadata definition, helping users find the right data sources while enforcing robust governance rules. These capabilities enable dynamic security policies that adapt based on user roles, session details or location. For example, you can consistently mask sensitive data such as Social Security numbers depending on a user's role or connection details. This streamlines data security management, ensuring compliance and reducing errors.

Data Products and Prepackaged Data Assets

Every agency has employees who have a deep understanding of the data they work with regularly. The Denodo Platform empowers these individuals to package and share this information so that their colleagues also can access and use it, including for AI applications.

This approach supports the scalability in the AI effort because packaged data products are readily reusable across a range of use cases. It also helps ensure data consistency and quality, which are key to building accurate and responsible AI applications.

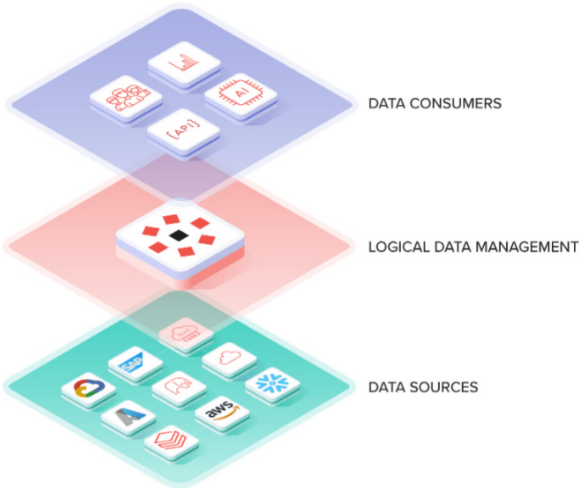
Intelligent Data Delivery

The Denodo Platform offers intelligent data delivery by leveraging AI and automation to enhance both data engineering and consumption.

- **For data engineers**, the platform provides AI-driven query optimization, automated recommendations and AI-generated SQL alongside intelligent tag and description suggestions, ensuring optimal data performance and organization.
- **For data consumers**, it simplifies data interaction by allowing natural language queries and delivering relevant answers directly within the platform. Additionally, it offers personalized data product recommendations based on past searches and usage, enabling more informed and efficient data use.

All this helps drive efficiency and accuracy as agencies look to build out their emerging AI implementations.

Intelligent Data Delivery at the Speed of Need



Accelerating AI Initiatives

How to enable GenAI success into the future with trusted, AI-ready data:

As government officials look to bring AI to the forefront, all these elements come together to support that effort. In simplifying and streamlining data scientists' work, this approach accelerates the path toward generative AI and other powerful use cases.

First, the logical data fabric supports retrieval-augmented generation (RAG) by providing a single access point for LLMs to gather data, no matter where it's stored or what format it's in. By sharing metadata about various data products, including their definitions and fields, LLMs can quickly identify the best sources to pull in the right context, ensuring accurate responses.

For example, when asked about the most common medical procedures among a specific group of patients, RAG helps guide the AI to return an answer that aligns with the agency's mission and use cases, rather than relying on generic information. By offering a unified and secure data access layer, the logical data fabric makes it easier to integrate enterprise data repositories into the RAG process.

With data context, the LLMs can go directly to the right data source, improving the accuracy of AI outputs and reducing the likelihood of hallucinations. Teams using the logical data fabric to build AI models get high-quality data, and they can use it to define rules and governance, ensuring that the data is used safely and responsibly.

The federal government is moving fast and hard to seize on the operational benefits that AI offers. With a logical data fabric, agencies can speed those efforts. They can break down silos and readily access the data they need to bring to life new AI-driven applications – and they can do it in a way that ensures consistent and accurate outcomes, with automation in support of strong data governance.

Learn more at www.denodo.com