**Notes about Denodo and LLMs**

**RAG’s relevance and reasons to implement this AI model**

Generative AI models struggle when you ask them about facts not covered in their training data. Retrieval Augmented Generation—or RAG—enhances an LLM’s available data by adding context from an external knowledge base, so it can answer accurately about proprietary content, recent information, and even live conversations. All that we have read about LLM applied to data management is related to applying this kind of updated IA in order to enhance the accuracy of the output.

The combination of retrieval and generation offers some benefits in data analysis:

* **Access to relevant and updated information:** instead of depending on a data historical model, RAG enables to get specific information recently stored in the database. This is ideal for reports that need fresh data such as finance analysis, market studies and changes in operational data.
* **Better responses:** by combining the retrieval of specific information with the ability to generate natural language responses, RAG provides more detailed and tailored answers to the query. This is useful when analyzing complex data and needing to generate automatic reports that answer specific questions from analysts or managers.
* **Automation of advanced queries:** in the context of data analysis, RAG can be used to automate the generation of insights based on natural language queries. For example, an analyst might ask questions like "What were the main reasons for the sales increase this quarter?" and the system could search relevant documents (such as sales reports, market analysis) and generate a detailed response.
* **Exploration of large data sets:** for companies that handle large volumes of documents or reports (such as system logs, call transcriptions, investigations, etc.), RAG allows for quickly finding the relevant information for analysis. This helps save time and reduces the effort required to manually search for relevant data before analyzing it.

**Why is LLM application important in Denodo?**

It significantly enhances how organizations can access, manage, and leverage their data for a variety of purposes, especially in the context of data virtualization and generative AI.

Denodo 9.0 key features regarding LLMs are the following:

* **Natural Language Queries**: Users can now interact with data using everyday language, removing the need for SQL knowledge. Denodo 9.0 integrates with platforms like **ChatGPT** and **AWS Bedrock**, allowing for seamless data queries through natural language.
* **AI-Powered Query Recommendations**: The platform uses AI to analyze user interactions and automatically recommends the best datasets, optimizing productivity and minimizing manual search efforts.
* **Generative AI Integration**: Denodo enables the use of **LLMs** (like those in Amazon Bedrock) with real-time, governed enterprise data, improving accuracy in generative AI applications.
* **RAG (Retrieval-Augmented Generation)**: This framework helps LLMs retrieve the most relevant, authoritative data from enterprise systems, ensuring accurate and secure responses in AI-driven applications.
* **Developer Productivity Enhancements**: Improved support for **DevOps**, task scheduling, and enhanced data-lake integration (Iceberg, Delta tables), making data lakes easier to manage and more accessible for analysis.

In summary, we could say that Denodo uses not only its own capacities of LLM but also integrates LLMs from other platforms (like **OpenAI**, **Amazon Bedrock** (for AWS) or **Google Cloud**). This enables them to offer “the best of both worlds”; natural language queries about their internal datasets and also the possibility of using more advanced AI models for predictive purposes.

Regarding Denodo's Partnership with Google Cloud, Denodo integrates with **Google Cloud’s Vertex AI** to accelerate innovation, combining Denodo's data virtualization with Google's LLM capabilities. This allows industries like **healthcare**, **finance**, and **education** to unlock new opportunities through AI, while maintaining a focus on ethical AI practices, privacy, and security.

[Demo: Accelerating Gen AI Applications with Denodo and Google Vertex AI](https://www.youtube.com/watch?v=tpQMPXYZRaA)

[Denodo Platform 8.0 Q3 2023 Update - AI-Generated queries](https://www.youtube.com/watch?v=w3AkHTQW3cs)

**Some extra information about Denodo**

LLMs can be used in various ways in data systems:

* **Natural Language Queries**: Users can interact with data by asking questions in plain language rather than writing SQL queries.
* **Data Summarization**: Automatically summarize complex data reports or trends, which can be useful for non-technical users or decision-makers.
* **Text Classification**: LLMs can classify unstructured text data (e.g., classify patient records based on diagnosis).

#### **How Denodo Integrates LLMs:**

Denodo has started exploring ways to integrate LLMs into its platform. This can enable natural language processing capabilities to interact with virtualized data more intuitively.

Some potential integration methods include:

1. **Natural Language Interface for Data Querying**: Using LLMs, Denodo can allow users to ask questions in natural language (e.g., “Show me the top 10 hospitals with the highest readmission rates”) instead of writing SQL queries. The LLM interprets the query and translates it into SQL.
2. **Data Enrichment**: LLMs can analyze unstructured data (e.g., doctor’s notes, patient feedback) and extract meaningful insights that can be integrated with structured data sources.
3. **Predictive Analytics**: By integrating with platforms like MindsDB, LLMs can assist in model training and predictive analytics, improving decision-making by understanding natural language input and context.

To start integrating LLMs into Denodo:

* **Research Denodo’s Latest Releases**: Denodo may release updates or modules specifically designed to interact with LLMs. Check Denodo’s **community forums** or official documentation for any extensions related to natural language queries or AI integrations.
* **Third-Party Integrations**: You can also explore external LLM providers (like OpenAI) that could integrate with Denodo’s API to provide natural language processing on top of your virtualized data.

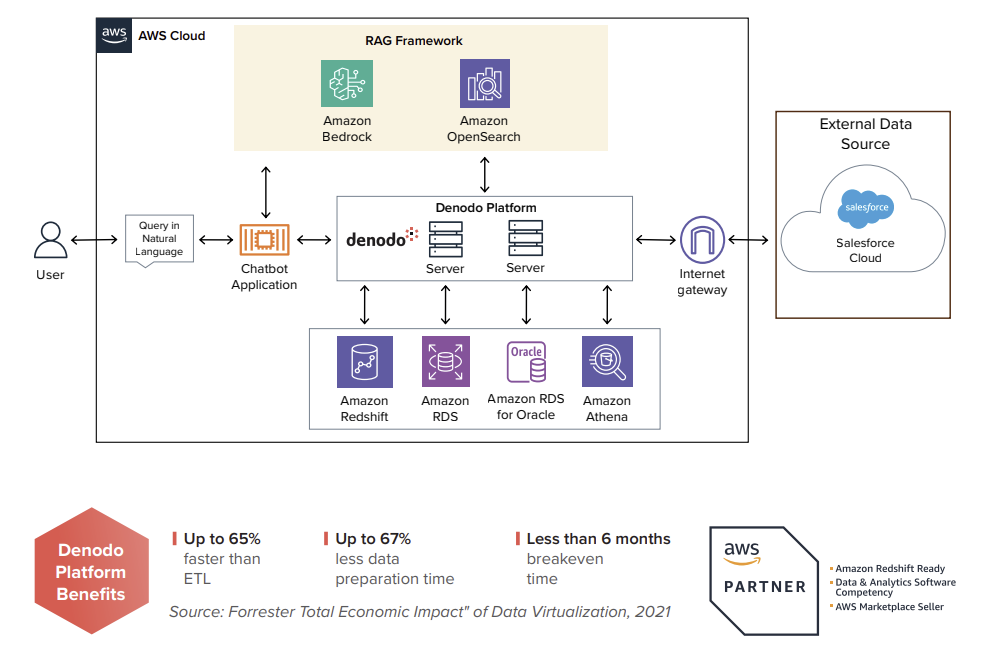
**DV-AWS Bedrock**

Despite the many promises of generative artificial intelligence (GenAI), many organizations are still struggling to build GenAI applications that meet business goals, and within a reasonable timeframe. Denodo and AWS are working to change that, by combining the Denodo Platform’s logical data management capabilities with Amazon Bedrock large language models (LLMs). This combination will streamline the development of powerful new GenAI enterprise applications that demonstrate security, privacy, and responsible AI.

Enterprise GenAI applications often fail to deliver their expected value because LLMs, “out of the box,” do not know where to find the most accurate, authoritative enterprise data which to answer questions. For enterprise applications, the right information is often not on the public Internet but on proprietary enterprise systems. The retrieval augmented generation (RAG) framework addresses this problem by introducing an intermediary step in the query process that points the LLM to the authoritative data sources, many of which will be internal to the enterprise.

**Key Benefits**

The Denodo Platform feeds Amazon Bedrock LLMs with governed, trusted, and AI-ready data from within the enterprise, within a RAG framework, leveraging Amazon OpenSearch as the vector database. The Denodo Platform enables a unified semantic layer above the underlying data sources, which simplifies access and facilitates end-to-end data governance and security. Leveraging a logical approach to data management, the Denodo Platform enables an enterprise-wide data fabric that can deliver real-time data to an LLM, minimizing reliance on extract, transform, and load (ETL) processes. With this architecture, all data in the organization can be accessed through RAG-enabled queries, yet data is only exposed to the authorized user who asks the question.



**Realizing the power of GenAI**

The joint Denodo/AWS solution enables a wide range of powerful use cases, including:

**1. Advanced Predictive Analytics for Business Intelligence**. The Denodo Platform’s data fabric integrates and harmonizes data across the enterprise, providing a unified view for analysis, and the Amazon Bedrock LLM can leverage this view to perform advanced predictive analytics, identify trends, and generate actionable insights for strategic decision-making.

**2. Risk Management and Compliance Monitoring**. Leveraging the Denodo Platform’s ability to consolidate and manage regulatory data, financial records, and operational data across the enterprise as a trusted data foundation, the Amazon Bedrock LLM can monitor compliance, detect anomalies, and predict potential risks, facilitating adherence to regulations.

**3. Enterprise Amazon-Bedrock-Enabled Chatbots.** Through RAG, the Amazon Bedrock LLM is directed to particular authoritative data sources within the enterprise when answering a question. The Denodo Platform provides enterprise data to the LLM that is trusted, secured, and authoritative, for the most accurate answers free of hallucinations or confabulations.

**4. Enhancing Developer Productivity and Simplicity.** Improved usability by end users via the adoption of natural language queries, increased productivity in development with an AI-based copilot that helps during the development of views, and simpler administration and stewardship of the system, for example with query performance tuning recommendations based on the past activity

**GenAI use case of a chatbot being used by a bank**

A user might ask the GenAI application a question, such as “How many loans were granted last week.” The LLM routes the question through the RAG framework and sends an SQL query to the Denodo Platform’s data fabric, which finds the data wherever it is stored, prepares it for the LLM, and delivers it back to the LLM, which puts the data into words and delivers them back to the user: “234 loans were granted last week.”

