The Denodo Platform

Enhancing Mission-Critical Federal Operations with Intelligent Data Delivery

Gaining timely, effective insights from data is crucial in today's fast-paced government environment. It's all about innovation, making smart decisions, and keeping up with changes. Imagine if everyone in your organization could instantly securely access and integrate data from across the organization. Unfortunately, data landscapes are constantly evolving, with increasing cloud workloads and data spread across more environments than ever. The rising demand for data, driven by technologies like generative AI, also challenges data teams to keep pace.

Departments and agencies are leveraging data lakes and cloud data warehouses to store the necessary data from across the organization. However, some data is likely to remain distributed due to extraction difficulties or data sovereignty laws, limiting the government's ability to react swiftly and innovate. A new approach is needed, one that enables a layer of intelligence above an organization's existing systems, to deliver data in the form that each use-case needs, at the speed of mission, regardless of where the data resides or what format it is in.

Proudly transforming data access and management across the U.S. Federal Government

Department of Defense

Denodo is reducing the time from data to decision, which is critical for leaders up and down the chain of command. Denodo has Authority to Operate (ATO) within the DoD and the platform is approved at DoD Impact Level 4, 5, and 6 (IL4, IL5, and IL6).

Civilian Agencies

Whether data is needed to improve operations, inform decision-making, or deliver better services more cost-effectively, Denodo has tremendous experience helping Civilian agencies transform data into the actionable insights necessary.

Intelligence Community

Accelerating information discovery and analysis is critical to maintaining intelligence advantage. Denodo can empower analysts with trusted AI-ready data from across data sources at the speed of need. And with an ATO in the Intelligence Community, we can ensure a secure deployment.

The Logical Approach to Data Management

The logical approach is based on a unified data delivery platform that simplifies how everyone in your organization accesses the various data systems. It's about stripping away the complexity and serving up the data in formats that make sense to the user, while accelerating delivery and adhering to data governance rules.

The Denodo Platform is the leading logical data management solution, enabling intelligent data delivery from distributed sources and presenting information in user-friendly formats for real-time, organization-wide access. It ensures the secure, timely delivery of high-quality data, empowering business users with seamless connectivity to a unified data source. Additionally, it enables security and compliance teams to monitor and manage all data access across all data sources, from a single point of control.

Benefits include:

- **Unified Data Access:** Breaking down data silos and offering a centralized virtual data abstraction layer for easy access to data, regardless of location or format

- **Data Tailored for Business Users:** Translating complex data into user-friendly formats, simplifying access, querying, and analysis without requiring technical skills or IT help

- **Simplified Data Discovery:** Enhancing visibility and accessibility with rich metadata, detailed lineage information, and AI-powered, intuitive search capabilities, to revolutionize data discovery and usage

- **Real-Time Insights:** Delivering real-time data, ensuring timely access for faster decision-making and a competitive edge

- **Enhanced Data Security and Compliance:** Ensuring integrity, security, and reliability using advanced features and third-party-tool integratio

- **Government Vetted:** Denodo has Authority to Operate (ATO) at select DoD and the intelligence community organizations. The platform is approved at DoD Impact Level 4, 5, and 6 (IL4, IL5, and IL6).
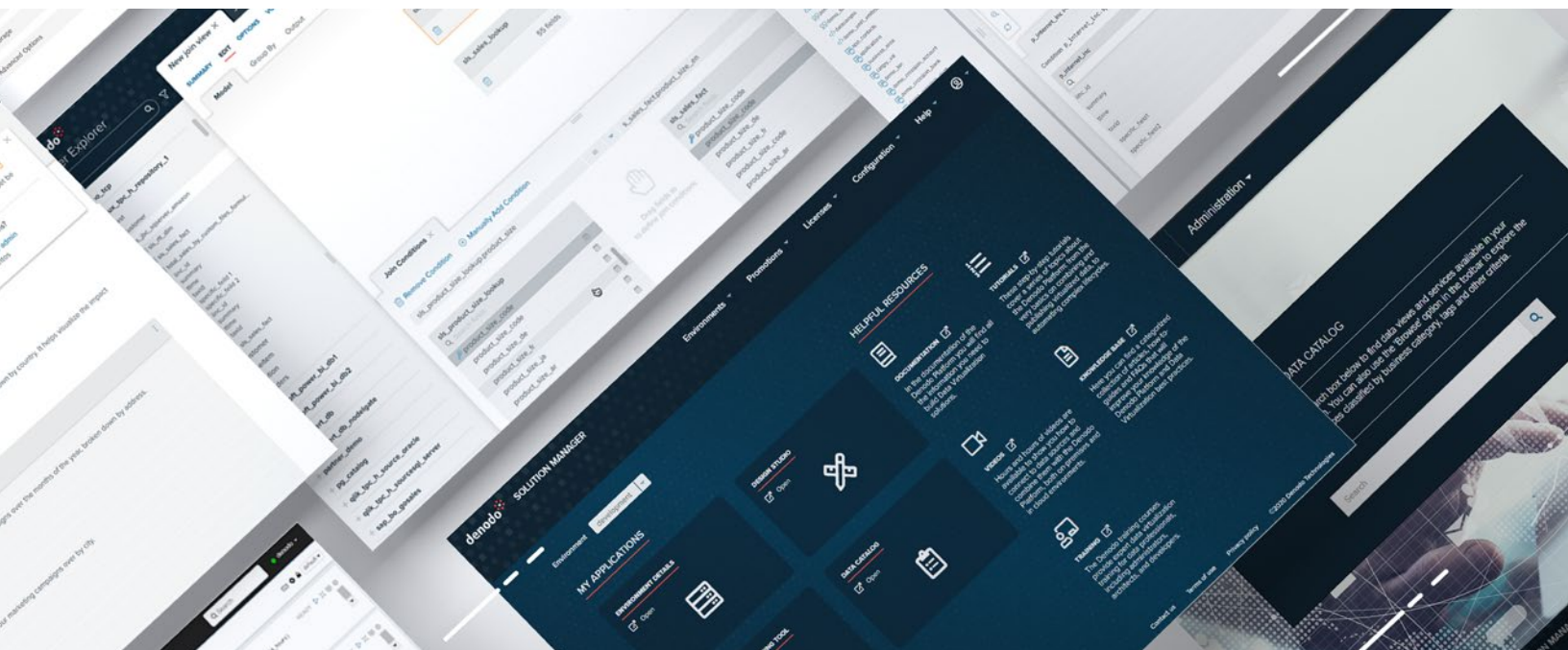
Denodo Platform Architecture

The Denodo Platform sits on top of an enterprise data landscape, enabling access to multiple, diverse data sources while making them appear as one "logical" data source to users. The Denodo Platform adheres to a logical data management architecture that is designed to optimize both traditional data sources (databases, enterprise data warehouses, data lakes, etc.) and other data sources (applications, files, web services, and the cloud) to meet a wide variety of use cases.
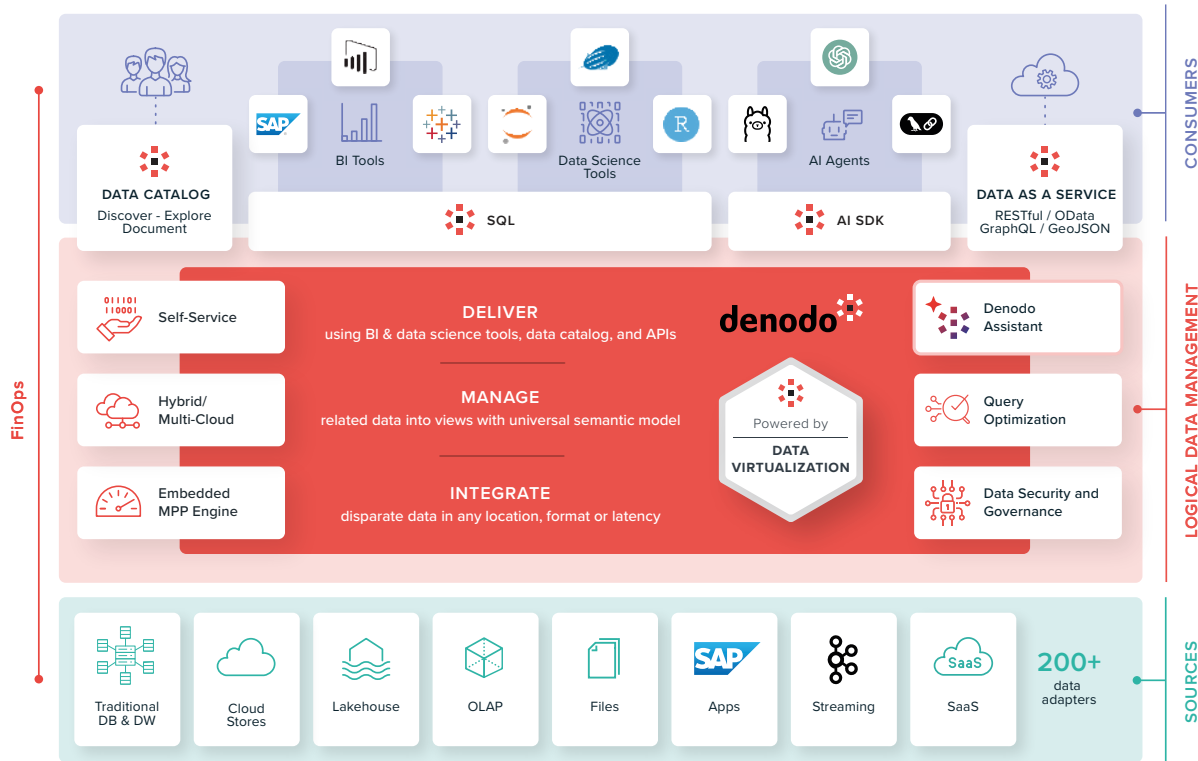

Denodo Platform Differentiators

- **Integrated Data Catalog:** Streamlines data access with automatic recommendations, enhanced collaboration, profiling, and natural language search powered by AI.

- **Universal Connectivity:** Simplifies connectivity to the widest range of data sources, with over 200 connectors to 3rd-party systems.

- **Embedded MPP Data Lake Engine:** Provides data consumers with a highly scalable and performant SQL engine, based on Presto, with which to access data lake data, while simplifying data access through an intuitive GUI interface.

- **A Unified Semantic Layer:** Facilitates data discovery, search and classification, and security and governance.

- **Global Security Policies:** Globally manage access control (masking, encryption, data restrictions, etc.) using security classifications, glossary terms, and tags.

- **Smart Query Acceleration:** Optimizes multi-source query execution through intelligent caching, cost-based optimization, and an AI-powered recommendations engine.

- **Financial Operations (FinOps):** Provides visibility into operating costs, for more effective management of cloud infrastructure costs.

- **Broad Delivery Options:** Provides more than JDBC or ODBC support for SQL queries by business users – developers can deliver data via a variety of options, including REST and GraphQL APIs, Kafka, and JMS message queues.

- **Data Preparation Wizard:** Empowers users of all technical backgrounds to easily customize datasets within the data catalog, enabling quick adaptation of data products for various use cases.

Denodo Assistant: Driving Efficiency and Better Outcomes

Denodo Assistant boosts business productivity by making data access and decision-making faster and smarter. Powered by the Denodo Platform's semantic layer and its extensive metadata, this intelligent assistant uniquely leverages technical, business, and operational metadata to enhance AI-driven data management tasks. By tapping into a unified and trusted data access layer, Denodo Assistant helps your teams work more efficiently, reducing the time and effort needed to find, understand, and access data. This intelligent assistant empowers technical and non-technical users, helping them make better decisions with guided recommendations, intuitive data access, and simplified workflows.
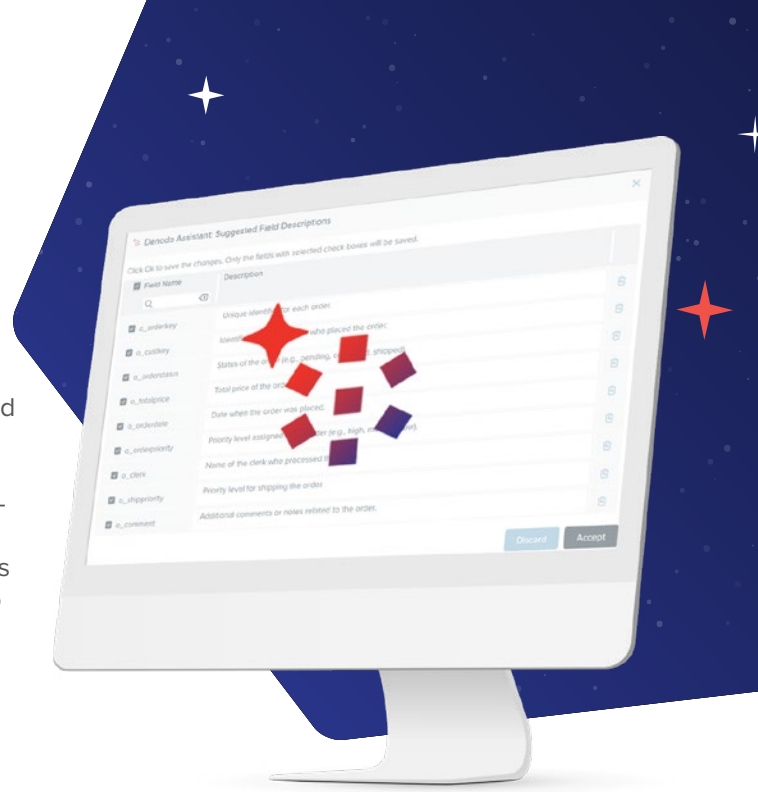
How Denodo Assistant benefits your organization:

- **Streamlined Data Access:** Enables teams to easily find and access the right data with natural language queries, eliminating the need for technical skills or IT support.

- **Data Product Recommendations:** It provides intelligent suggestions based on past searches, helping users quickly identify the most relevant data for their needs and leading to faster, more informed decisions.

- **Optimized Data Preparation:** Simplifies data preparation with AI-driven recommendations, enabling non-technical users to easily tailor data to their needs.

- **Simplified Data Descriptions:** This feature automates the addition of descriptions to data views and columns, enhancing understanding and making it easier for teams to use the data effectively.

- **Automatic Summary Recommendations:** Leverages machine learning and active metadata to suggest summary caches, optimizing query performance and efficiency according to business rules.

- **Query Wizard Recommendations:** Guides users through query creation, enabling all skill levels to generate accurate queries with ease.

- **Intelligent Autocompletion in VQL Shell:** Provides context-aware query suggestions, reducing errors and speeding up query writing.

- **LLM-Based Data Enrichment Function:** Enables users to automatically summarize text, extract information, analyze sentiment, remove sensitive data, and translate text using large language models (LLMs), available as a function call in VQL.

Denodo Enterprise Plus Capabilities Sheet

DATA SOURCES

Relational Databases
- Generic (JDBC)
- IBM DB2 8.2, 9, 10, 11, and higher; 9,10, and 11 for z/OS
- Denodo Virtual DataPort - 8.0, 9
- Apache Derby 10 and higher
- Informix 7, 12
- Microsoft SQL Server 2000, 2005, 2008, 2008 R2, 2012, 2014, 2016, 2017, 2019, and higher
- MySQL: 4, 5, 8, and higher
- Oracle: 8i, 9i, 10g, 11g, 12c, 12c In-Memory, E-Business Suite 12, 18c, 19c, and higher
- PostgreSQL 8, 9,10,11,12, and higher
- Sybase ASE / SAP ASE 12, 15
- MS Access

In-Memory Databases
- SAP HANA 1.0, 2.0
- Oracle TimesTen 11g
- Oracle 12c In-Memory

Parallel Databases and Appliances
- ClickHouse
- Exasol 7.1
- Greenplum 4.2
- Vertica 7, 9
- Netezza 4.6, 5.0, 6.0, 7.0, and higher
- Oracle Exadata
- ParAccel 8.0.2 (by using ParAccel 2.5.0.0 JDBC3g with SSL driver)
- SybaseIQ
- SQreamDB
- Teradata 12, 13, 14, 15, 16, 17
- Yellowbrick

Cloud Data Warehouse / RDBMS
- Alibaba ApsaraDB for OceanBase MySQL
- Alibaba ApsaraDB for OceanBase Oracle
- Alibaba ApsaraDB RDS for MySQL
- Alibaba ApsaraDB RDS for PostgreSQL
- Alibaba ApsaraDB RDS for Microsoft SQL Server
- Alibaba ApsaraDB PolarDB for MySQL
- Alibaba ApsaraDB PolarDB for PostgreSQL
- Alibaba ApsaraDB AnalyticDB for MySQL
- Alibaba ApsaraDB AnalyticDB for PostgreSQL
- Alibaba MaxCompute
- Amazon Redshift
- Amazon Athena
- Amazon Aurora (MySQL and PostgreSQL)
- Amazon DynamoDB
- Azure Cosmos DB
- Azure SQL Database
- Azure Synapse SQL (aka Azure SQL Data Warehouse)
- Delta Lake
- Google AlloyDB
- Google BigQuery
- GCP (Google Cloud Platform) SQL for MySQL; GCP SQL for PostgreSQL
- Google Spanner (Includes support for the service Spanner Data Boost)

- MongoDB Atlas
- Snowflake

Big Data
- Hive 0.13.0 (Hive Server 2); 1.1.0 (Hive Server 2); 2.0.0 (Hive Server 2); 3.1.2 and higher (Hive Server 3); Hive for Cloudera 1.1.0; Hive for Hortonworks 1.2.1
- Impala 1.2.4., 2.3, 3.x kudu
- Spark SQL 1.5, 1.6, 2.x, 3.x and higher
- PrestoDB
- Trino
- Databricks

NoSQL
- MongoDB
- Cassandra 3.x

Multi-Dimensional Sources
- Multidimensional database (generic)
- Azure Analysis Services
- SAP BI 3.x and BW 7.x
- Mondrian: 3.x
- IBM Cognos TM1
- MS SQL Server Analysis Services
- Essbase

Data Lake Storage/Formats
- S3
- Azure Data Lake Storage Azure Blob Storage
- Google Cloud Storage
- Parquet/Avro
- Delta/Iceberg

Web Services
- SOAP
- REST (XML, RSS, ATOM, JSON)
- OData v2.0 and v4.0

Flat and Binary Files
- CSV, pipe-delimited, regular expression-parsed
- MS Excel xls 97-2003
- MS Excel xlsx 2007 or later
- MS Access
- XML
- JSON
- SAS Files (SAS7BDAT)
- All files can be local or in remote filesystems, through FTP/ SFTP/FTPS, and in clear, zipped and/or encrypted format.

Indexes and unstructured content
- CMS, file systems, text
- Elasticsearch

Cloud, SaaS, Web Sources
- Adobe Analytics
- AWS
- Google Analytics
- Google Sheets
- Facebook
- LinkedIn

- MS Azure Data Lake
- MS Sharepoint
- MS Dynamics 365 Business Central / Customer Engagement
- Marketo
- ServiceNow
- Salesforce
- Twitter
- Workday
- many more through configurable JSON and XML adapters

Active Directory as Source or Leveraging Security
- LDAP v3
- Microsoft Active Directory 2003, 2008

Streaming/Messaging systems
- Kafka
- MQSeries
- SonicMQ
- ActiveMQ
- Tibco EMS
- Other JMS compatible services

Semantic Repositories
- Semantic repositories in Triple Stores / RDF accessed through SPARQL endpoints.

Packaged Applications
- SAP ERP/ECC (BAPIS and tables)
- Oracle E-Business Suite 12
- Siebel
- SAS 9.4

Hierarchical Databases
- Adabas (SOA Gateway and Denodo's SOAP connector): 5, 6

Denodo SDK for Custom Connectors

PUBLISHING OPTIONS

- SQL Based access via JDBC, ODBC and ADO.NET
- Web Services
  - REST
  - OData
  - Open API (a.k.a Swagger)
  - GraphQL
  - SOAP
- OAuth, OAuth 2.0 (JWT)
- SAML
- SSL
- WS-Security
- JMS listeners for message queues
- Denodo Scheduler for batch process and lite ETL

DATA CATALOG E E+

Cataloging
- Web-UI for seamless data discovery and exploration for business users
- Descriptions, documentation, and custom properties

- Business categories and tags Intelligent search with smart ranking of results

Governance
- Graphical data lineage
- Integrated request management (access, changes, data quality issues, etc.)
- Endorsement of datasets, comments, warnings, etc.
- Usage statistics: who uses what data, when and how
- Data profiling information

Self-Service
- Last-mile data preparation wizards for customizing datasets by non-technical users
- Full-featured SQL shell facilitates the execution of complex queries
- Export to CSV, Excel, and Tableau Data Extracts
- Save personal queries for easy access
- Query sharing features

PERFORMANCE OPTIMIZATIONS

- Smart Query Acceleration for Analytics E E+
  - Aggregate Aware Summaries
- Massively Parallel Processing (MPP) integration E E+ for Query Acceleration and Caching
- Full and partial aggregation and join pushdown, even in federated views
- Support for alternative data sources
- On-the-fly data movement for optimization
- Cost-based optimization (data statistics, data source indexes, data source execution model and parameters, network transfer rates)
- Pushdown of selections/projections/joins/groupby operations also on federated views
- Multiple join strategies
- Simplifying partitioned unions (Partition pruning)
- and many more

CACHE AND DATA MOVEMENT OPTIONS

- Multi-mode caching: full, partial, incremental, or total refresh, event-based or scheduled, configured at the view level, incremental queries for SaaS sources
- AlloyDB for PostgreSQLAmazon Athena
- Amazon Aurora MySQL
- Amazon Aurora PostgreSQL
- Amazon Redshift
- Azure SQL
- Azure Synapse SQL (previously known as Azure SQL Data Warehouse)
- Databricks
- Exasol
- Google BigQuery
- GCP Cloud SQL for MySQL
- GCP Cloud SQL for PostgreSQL
- Hive 2.0.0; Hive 3.1.2, and higher (HiveServer2)
- IBM DB2 (8, 9, 10, 11, and higher for LUW; 9,10,11 for z/OS)
- Impala 2.3; 3.x Kudu
- MS SQL Server (2000, 2005, 2008, 2008R2, 2012, 2014, 2016, 2017, 2019, and higher)
- MySQL 4, 5, 8, and higher

- Netezza 6, 7, and higher
- Oracle 8i, 9i, 10g, 11g, 12c, 12c In-Memory, 18c, 19c, and higher
- Oracle TimesTen 11g
- PostgreSQL 9, 10, 11, 12, and higher
- Presto
- SAP HANA 1 and 2
- Snowflake
- Spark SQL 2.x, 3.x, and higher
- Teradata 12, 13, 14, 15, 16, and 17
- Trino 4xx
- Vertica 7 and 9
- Yellowbrick
- Configurable "generic" adapter for other databases with JDBC drivers

DATA PIPELINES

- Remote Tables (created through UI or stored procedure)
- Denodo Scheduler
- VQL stored procedures

EMBEDDED MPP E+

- MPP engine based on Presto to accelerate access to data lake
- Graphical introspection of object storage (S3, ABFS, GFS, HDFS, etc.)
- Support for Parquet, Delta, and Iceberg
- New advanced optimization techniques to federate data lake content with any other data source

THIRD-PARTY MPP OPTIONS

- Impala
- Presto
- Spark 1.5, 1.6, 2.x
- Databricks 2.x

DATA GOVERNANCE

- Data source refresh, change impact analysis, dependency tree, full data lineage
- Denodo Governance Bridge: integration with IBM Information Governance Catalog
- API to publish metadata and lineage information to data governance tools like Informatica EDC, Collibra, etc.

SECURITY

Data in Motion – secure channels
- Using SSL/TLS
- Client-to-Denodo and Denodo-to-source
- Available for all protocols (JDBC, ODBC, ADO.NET and WS)

Data at Rest - secure storage
- Cache: third party database. Can leverage its own encryption mechanism
- Swapping to disk: serialized temporarily stored in a configurable folder that can be encrypted by the OS

Encryption/Decryption
- Support for custom decryption for files and web services
- Transparent integration with RDBMs encryption
- Encrypted metadata import/export

User and Role Based Including Integration with AD/LDAP
- Row and column level authorization
- Advanced customizable masking
- Custom policies for specific security constraints and integration with external policy servers

Global Policies
- Tag-based security policies
- Support for RBAC and ABAC
- Column and row restrictions, multiple masking options, deny execution

Authentication
- Native and LDAP/Active Directory based Support for Kerberos and Windows SSO
- Kerberos
- NTLM
- OAuth, OAuth 2.0 (JWT)
- SAML
- Two-factor authentication (through supported identity providers: Okta, Duo, etc.)
- SSL
- WS-Security
- Pass-through session credentials to leverage existing source privileges

ADVANCED SEMANTICS E+

- Global security policies
- Tagging at column level
- Support for the importing of external data governance tags from Collibra and other external data catalogs

DATA MODELING

- Design Studio: Web-based development studio for data modeling
- Desktop version also available
- Bottom-Up and Top-Down (through Interface Views)
- Integration with third-party modeling tool E E+
  - ER/Studio Data Architect
  - ERwin Data Modeler
  - IBM InfoSphere Data Architect
  - SAP PowerDesigner
  - Sparx Systems Enterprise Architect

DATA QUALITY

- Library of transformation, filter, and matching functions and quality rules for validating, cleansing, enriching, standardizing, matching, and merging data
- Extensible through custom functions
- Integration with external data quality tools

MONITORING

- Denodo Diagnostic and Monitoring Tool **E** **E+** (DMT) integrated in the Solution Manager
- Extensible usage and metadata dashboards integrated in Apache Superset
- FinOps dashboard to monitor and understand key metrics associated with cloud costs like egress, query cost, etc.
- Detailed monitoring information is available in logs for integration with log management tools like Splunk, ELK, Cloudwatch, etc.
- Monitoring is also available via SNMP and JMX standards. Therefore interoperate with most leading Systems Management packages (e.g., HP OpenView, Nagios, Zenoss, Osmius, IBM Tivoli and Microsoft WinRM)

DENODO ASSISTANT

Data Catalog
- Natural language queries with GenAI
- Support for OpenAI, Azure OpenAI, AWS Bedrock, and custom LLMs.
- Personalized dataset recommendations
- Smart SQL-fragment autocomplete based on previous activity

Design Studio
- Summary cache recommendations for Smart Query Acceleration
- View and column description suggestions
- Business friendly column name suggestions
- Diagnose incorrect queries and explain queries

VQL Function
- Automatically summarize, extract, analyze sentiment, remove sensitive data, and translate text using LLMs.

OPERATIONS

- Solutions Manager to automate operations and promotions tasks
  - Centralized management and distribution of updates to clients
  - Centralized management of license keys
  - Define promotion revisions and their dependencies and deploy them to a production cluster with zero downtime
  - Centralized management of data source properties and logs
  - REST API for automation of tasks from DevOps tools (e.g. jenkins)
- Integrated Infrastructure Management for Cloud (AWS)
  - Creation and management of clusters: define type of EC2 instances, number of EC2 instances, etc.

- Creation of load balancers and Auto Scaling groups.
- Installation and launch of Denodo servers.
- Update the Denodo version.
- Enable SSL in the Denodo servers.
- Multi-User Development with Version Control integration
  - Git
- Resource Manager to limit and allocate resources to each session, role, or user in a way that optimizes resources utilization for each application
  - Change resources priority
  - Enforce limited timeouts or limits on number of rows
  - Add daily quotas per minute/day/month: e.g. only 50 queries per day

DEPLOYMENT PATTERNS

- On-premises, private cloud, public cloud
  - On-premises, private cloud, public cloud
  - Basic single server configuration
  - HA cluster with load balancing (Active-Passive and Active-Active)
  - Shared or distributed local cache
  - Geographically distributed server environments
  - Multiple Denodo instances, peer-to-peer or multilayered
  - Containerization support through Docker
- Public cloud
  - Denodo Platform for AWS
  - Denodo Platform for Azure
  - Denodo Platform for GCP
  - Denodo Platform for Alibaba Cloud
- Auto-scaling support both in AWS and Azure
- Agora - the Denodo Cloud Service

OPERATING SYSTEMS

- **Microsoft:** Windows Server 2022, Windows Server 2019, Windows Server 2016, Windows 11, and Windows 10
- **Linux:** Amazon Linux 2023, Ubuntu 20.04 LTS or later
- CentOS 9.x and 8.x, Red Hat Enterprise Linux (RHEL) 9.x and 8.x, Oracle Linux 9.x and 8.x, SUSE Linux Enterprise 15.x

MINIMUM HARDWARE REQUIREMENTS

- Processor: Intel Xeon quad-core or similar. High-load scenarios or cases with complex calculations may require 8 cores or more.
- Physical memory (RAM): 16 gigabytes of memory so the Denodo server can allocate a runtime heap space up to 8 gigabytes.
- Disk space: Minimum: 5 gigabytes, Recommended: 100 gigabytes. Denodo only needs around 1 GB of disk space. If the cache is installed on the same server, more disk space will be required.