The Denodo Platform

Enabling Intelligent Data Delivery

Gaining timely, effective insights from data is crucial in today's fast-paced business environment. It's all about innovation, making smart decisions, and keeping up with market changes. Imagine if everyone in your company could instantly 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.

Traditional data management approaches involve moving and merging data into a central repository, such as a data warehouse or lake. This process can be expensive, time-consuming, and requires further preparation for various use cases, ultimately slowing business innovation and responsiveness. A new approach is needed—one that complements traditional methods by adding a centralized data access layer atop all data repositories. This layer simplifies complexity, presents data in business-friendly formats, and accelerates delivery while adhering to predefined semantics and governance rules.

Data Self-Service

Landsbankinn streamlined its data processes and enhanced accessibility, empowering over 80% of targeted users, including those with limited data skills, to make informed decisions.

Data Foundation for Improved Customer Experience

LeasePlan improved customer satisfaction by creating new programs like predictive vehicle maintenance, incorporating both internal and external data sources.

The Logical Approach to Data Management

The logical approach is based on a unified data delivery platform that simplifies how everyone in your business 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 integration
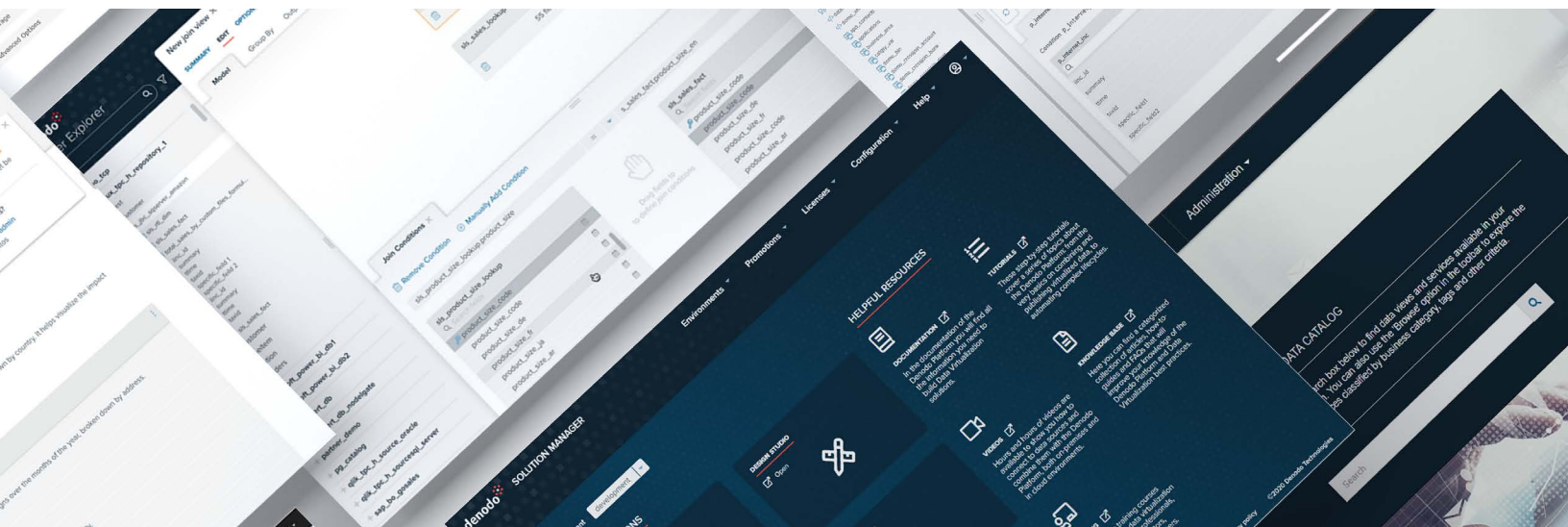
SUNBELT RENTALS

IT Infrastructure Modernization

Sunbelt Rentals accelerated access to diverse data sources, boosting productivity by 200% by retrieving data from diverse systems, including over 500 ERP-system tables in under three minutes, without engineering support.

CITY FURNITURE

Enhance Operational Efficiency, Agility and Resilience

City Furniture enhanced operations and profitability by developing a single source of truth across sales, supply chain, merchandising, and operations, resulting in improved market timing, expanded market share, and increased profits.

Seacoast Bank

Centralized Governance, Risk, and Compliance

Seacoast Bank improved security and compliance efforts by using data for business process automation and making information readily available to key managers, including risk managers.

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 and 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.

Agora
THE DENODO CLOUD SERVICE

Agora has all of the robust capabilities of the Denodo Platform and offers them as a fully managed, cloud-based solution that simplifies data management. It provides seamless, flexible deployment options without compromising the platform's simplicity and power. Agora enables organizations to offload infrastructure and operations management, enabling IT and operational teams to focus on business objectives and leverage the full potential of their data. With Agora, organizations can maintain agility in a dynamic business environment.

COMPARING THE VARIOUS SUBSCRIPTION TIERS

There are several flexible subscription options to choose from, designed to suit the needs of various projects, from small departmental projects to enterprise-wide digital transformation.

	P DENODO PROFESSIONAL — Small, single-use-case projects within individual departments	**S** DENODO STANDARD — Multiple use cases within individual departments	**E** DENODO ENTERPRISE — Enterprise-wide deployment for multiple use cases and groups and large data volumes	**E+** DENODO ENTERPRISE PLUS — Comprehensive automation, collaboration, and advanced security for enterprise-wide deployments
Number of Data Sources Supported	5	Unlimited	Unlimited	Unlimited
FinOps Logging and Integrated Dashboard	✔	✔	✔	✔
Available as a SaaS		✔	✔	✔
Integration with Version Control Systems (VCS)		✔	✔	✔
Smart Query Acceleration using summaries			✔	✔
VQL procedures			✔	✔
Advanced Diagnostic & Monitoring Tool			✔	✔
Data Catalog			✔	✔
Data Catalog: Dataset collaboration through endorsements, warnings, and deprecation notes			✔	✔
Integration with external Massive Parallel Processing (MPP) engines like Impala, Spark, and others			✔	✔
Integration with 3rd party data modeling tools (ER/Studio, Erwin, etc.)			✔	✔
Embedded Presto-based MPP engine				✔
Automatic recommendation of summaries				✔
Global security policies				✔
Import data governance tags from external catalogs				✔
Data Catalog: Natural language queries and AI-based recommendations				✔

Capabilities

DATA SOURCES

Relational Databases
- Generic (JDBC)
- IBM DB2 (JDBC): 8, 9, 10, 11, 12 for LUW, 9,10 for z/OS, AS400
- Multi Layered Denodo deployments (JDBC): 5.5, 6.0, 7.0, 8.0
- Apache Derby (JDBC): 10
- Informix (JDBC): 7, 12
- MS SQL*Server (JDBC, ODBC): 2000, 2005, 2008, 2008R2, 2012, 2014, 2016, 2017
- MySQL (JDBC): 4, 5, 8
- Oracle (JDBC): 8i, 9i, 10g, 11g, 12c, 18c, 19c
- Oracle E-Business Suite (JDBC): 12
- PostgreSQL (JDBC): 8, 9, 10, 11, 12
- Sybase Adaptive Server Enterprise (JDBC): 12, 15
- MS Access (ODBC)

In-Memory Databases
- SAP HANA (JDBC): 1,2
- Oracle TimesTen (JDBC): 11g
- Oracle 12c In-Memory

Parallel Databases and Appliances
- Exasol (JDBC)
- GreenPlum (JDBC): 4.2
- HP Vertica (JDBC): 7, 9
- Netezza (JDBC): 4.6, 5.0, 6.0, 7.0
- Oracle Exadata (JDBC): X5-2
- ParAccel 8.0.2 (by using ParAccel 2.5.0.0 JDBC3g with SSL driver)
- SybaseIQ (JDBC) 12.x, 15.x
- Teradata (JDBC): 12, 13, 14, 15, 16, 17
- Yellowbrick

Cloud Data Warehouse / RDBMS
- Alibaba ApsaraDB for OceanBase MySQL (JDBC)
- Alibaba ApsaraDB for OceanBase Oracle (JDBC)
- Alibaba ApsaraDB RDS for MySQL (JDBC)
- Alibaba ApsaraDB RDS for PostgreSQL (JDBC)
- Alibaba ApsaraDB RDS for Microsoft SQL Server (JDBC)
- Alibaba ApsaraDB PolarDB for MySQL (JDBC)
- Alibaba ApsaraDB PolarDB for PostgreSQL (JDBC)
- Alibaba ApsaraDB AnalyticDB for MySQL (JDBC)
- Alibaba ApsaraDB AnalyticDB for PostgreSQL (JDBC)
- Amazon Redshift (JDBC)
- Amazon Athena (JDBC)
- Amazon Aurora (JDBC)
- Amazon DynamoDB
- Azure Cosmos DB
- Azure SQL Database
- Azure Synapse Analytics
- Delta Lake
- Google AlloyDB (JDBC)
- Google Big Query (JDBC)
- Google Cloud SQL (JDBC)
- Google Spanner DatBoost (JDBC)

- MongoDB Atlas
- Snowflake (JDBC)

Big Data
- Apache Hive (JDBC): 0.12, 1.1.0, 1.1.0 for Cloudera, 1.2.1 for Hortonworks, 2.0.0
- Impala (JDBC): 2.3
- Spark SQL (JDBC): 1.5, 1.6, 2.x, 3.x
- PrestoDB (JDBC)
- PrestoSQL / Trino (JDBC)
- Databricks Delta 2.x

NoSQL
- MongoDB
- Cassandra
- HBase

Multi-Dimensional Sources
- SAP BW (BAPI/XMLA): 3.x
- SAP BI 7.x (BAPI): 7.x
- Mondrian (XMLA): 3.x
- IBM Cognos TM1
- MS SQL Server Analysis Services 200x
- Essbase (XMLA): 9, 111

Data Lake Storage
- S3
- Azure Data Lake Storage
- Azure Data Lake Storage Gen 2
- Azure Blob Storage
- Google Cloud Storage
- Parquet
- Avro

Web Services
- SOAP
- REST (XML, RSS, ATOM, JSON)
- OData

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 6.4, 6.7

Cloud, SaaS, Web Sources with Simplified OAuth Security
- Adobe Analytics
- Amazon
- Google

- Google Sheets
- Facebook
- LinkedIn
- MS Azure Data Lake
- MS Sharepoint (via OData)
- MS Dynamics 365 Business Central / Customer Engagement
- Marketo
- ServiceNow
- Salesforce (SOQL)
- Twitter
- Workday
- many more through configurable JSON and XML adapters

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

Message Queues
- 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 (SAS JDBC Driver): 7 and higher

Mainframe
- IMS
    - IBM IMS native drivers: 8, 9
    - IMS Universal Drivers: 11

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
- Massive 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
- Amazon Athena
- Azure SQL
- Azure SQL Data Warehouse
- Azure Synapse Analytics
- Amazon Redshift
- Databricks 2.x
- Delta Lake
- IBM DB2 (8, 9, 10, 11 for LUW, 9,10,11 for z/OS)
- Hive 2.0.0
- Impala
- MS SQL Server (2000, 2005, 2008, 2008R2, 2012,

2014, 2016, 2017)
- MySQL (4 and 5)
- Netezza (6 and 7)
- Oracle (8i, 9i, 10g, 11g, 12c, 12c in-memory, 18c, 19c)
- Oracle TimesTen 11g
- PostgreSQL (9 and 10)
- Presto
- SAP HANA
- Snowflake
- Spark (1.5,1.6 and 2.x)
- Teradata (12, 13, 14, 15, and 16)
- 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, ADFS, 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
- Base64
- 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

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 DQ 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)

EMBEDDED AI E+

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

Design Studio
- Performance recommendations for Smart Query Acceleration using usage history

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 the Denodo servers.
    - Update of Denodo version
    - Enable SSL in the Denodo servers.
- Multi-User Development with Version Control integration

- Subversion
- Microsoft TFS
- 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 multi-layered
    - 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

OPERATING SYSTEMS

- Microsoft Windows (32-bit and 64-bit platforms): Windows Server 2019, Windows Server 2016, Windows Server 2012, Windows Server 2008, Windows 10, Windows 8.1, and Windows 7
- Linux (32-bit and 64-bit platforms): Red Hat Enterprise Linux (RHEL) 6 and 7, Oracle Linux 6 and 7, Ubuntu 12.04 LTS and 14.04 LTS, CentOS 6 and 7
- Any Java 11 or greater compatible OS

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..

denodo