Denodo Platform 8.0 Datasheet

Today's data management landscape is becoming increasingly complex, as data is spread out across many heterogeneous data systems (data warehouses, columnar databases, MPP engines, specialized data stores, SaaS/cloud applications, etc.) that can be in multiple locations (on-premises, cloud, etc.). This makes it difficult to offer a unified view of the data to business applications, and to guarantee that governance policies and rules are enforced across the data delivery chain.

Logical data fabric is a vision of a unified data delivery platform that abstracts access to multiple data systems for business consumers, hiding the complexity and exposing the data in business friendly formats, while at the same time guaranteeing the delivery of data according to predefined semantics and governance rules, solving the main issues described above.

Data virtualization is the key technology to achieve the logical data fabric vision. As the only data integration style designed for distributed architectures, data virtualization provides a logical data access layer on top of multiple heterogeneous systems in hybrid, distributed architectures. Today, the data management ecosystem is distributed in nature, so a logical data access layer such as data virtualization is the best fit.

Denodo Platform 8.0 is a significant step forward to achieve the vision of a logical data fabric. It incorporates features that accelerate the delivery of governed data to business applications, in the most appropriate format for each consumer, across multiple, geographically distributed, heterogeneous systems.

Denodo Platform 8.0 offers stronger support for core data virtualization use cases: logical analytical architectures, logical data warehouses, and data services APIs. But it goes beyond traditional data virtualization scenarios to better support new types of users and new types of use cases, such as data science and machine learning (ML) initiatives.

In addition, it is a big step forward in platform as a service (PaaS) cloud strategy, with capabilities for automatically managing the cloud infrastructure from a centralized web console.

What Are the Key Features of Denodo Platform 8.0?

Denodo Platform 8.0 offers an enhanced, unified user experience with a full web-based interface for all Denodo Platform components, making it much easier to manage the Denodo Platform in hybrid on-premises/cloud deployments.

The new Denodo Central Web Console, integrated in the Solution Manager, provides a single entry point to all Denodo tools, for all users, in all Denodo Platform environments, both on-premises and in the cloud. It supports SSO using Kerberos, SAML, OpenID and OAuth, enabling seamless connectivity to all Denodo Platform tools. It also provides tighter integration between tools (e.g., the Diagnostic and Monitoring Tool is now integrated in the Solution Manager).

Developers now benefit from a new Web-based Studio tool with which to develop views and data services (They can still use the desktop version, which is still supported in Denodo Platform 8.0). This new tool has been designed with the principle of maintaining and enhancing the Denodo Platform's traditional ease of use.

In the performance area, Denodo Platform 8.0 introduces a new concept, smart query acceleration, which accelerates query execution in logical data warehouse / data fabric architecture. Partial aggregates of fact and dimension tables (called "summaries"), which are commonly joined together in many queries of a certain type, are precomputed and used to accelerate the execution of future queries. This technique achieves significant performance gains and facilitates the building of ad-hoc queries in a self-service scenario, as these performance enhancements are handled behind the scenes by the Denodo Platform. This extends the Denodo Platform's leadership in distributed query optimization, adding to the strength of other features such as advanced query pushdown, MPP acceleration, and sophisticated caching mechanisms.
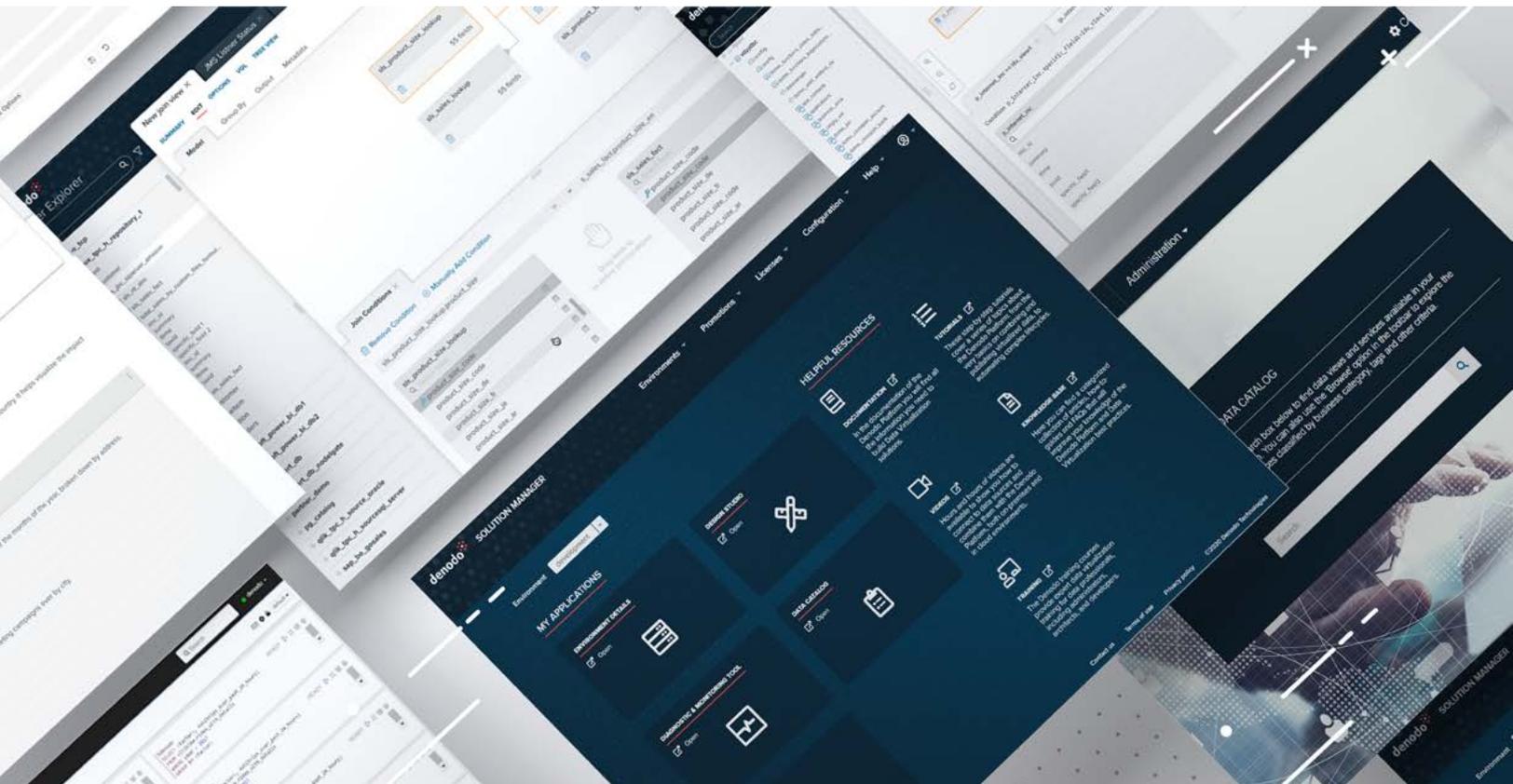
One of the most significant features of Denodo 8.0 is automated infrastructure management in the cloud, which automates all of the tasks related to installing, configuring, deploying, and upgrading Denodo Platform clusters. Initially available for AWS (Azure support will follow soon), this functionality is delivered through the new Solution Manager, which now provides a web-based UI with which to define and configure clusters, indicating user preferences in aspects such as TLS configuration, load-balancing, autoscaling, etc. Once the clusters have been defined using this UI, users can simply press "Start," and the clusters will be automatically installed and created. This UI also provides integrated monitoring.

Denodo Platform 8.0 offers advanced support for data services with flexible delivery options (REST, SOAP, OData, OpenAPI for documentation), the ability to expose data in multiple formats (XML, JSON, HTML, RSS) and support for the latest security protocols (OAuth, JSON Web Tokens, SAML, Kerberos, HTTPS, HTTP Basic Digest Authentication, or WS-Security). In this area, Denodo Platform 8.0 also adds support for GraphQL, one of the fastest growing data services standards. GraphQL is a query language for APIs. It simplifies the querying of multiple REST endpoints, and with Denodo Platform 8.0, this is achieved with zero code, and with much better performance, as it leverages the Denodo advanced query optimization engine.

To further facilitate data scientists' access to data through the Denodo Platform, Denodo Platform 8.0 introduces a new tool, the Denodo Notebook, which is based on Apache Zeppelin. The Denodo Notebook is fully integrated with the Denodo security system including SSO support, so all security and data governance policies defined at the Denodo Platform layer are enforced when data scientists use the Denodo Notebook.

For business users, the Denodo Data Catalog includes a new design that has the same look-and-feel as the other Denodo Platform tools, to achieve a more integrated experience. Now, the catalog includes new features such as automatic recommendations; enhanced collaboration with endorsement, comments, etc.; and enhanced profiling and search features (smart ranking) .
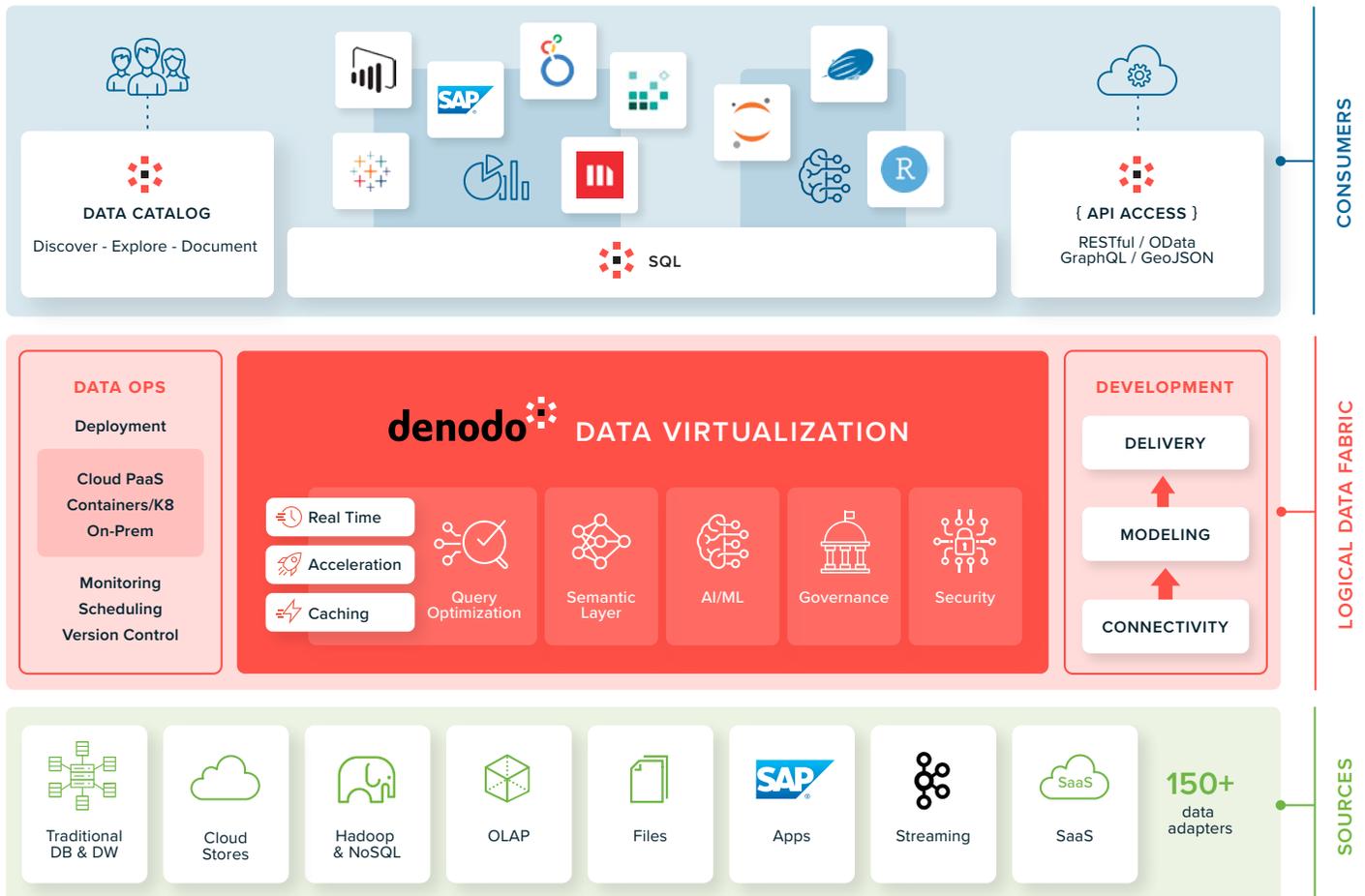

Fig. 1 Denodo Platform 8.0 Architecture

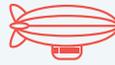To summarize, the key features of Denodo Platform 8.0 are:

A full web-based interface for all Denodo tools with SSO support: an integrated, web-based experience across all tools.

Web-based Design Studio tool for developers, providing ease-of-use across all of the steps in the data-service development process.

Smart query acceleration for analytics: Partial aggregates ("summaries") are pre-computed to accelerate future queries. Denodo Platform 8.0 provides this acceleration mechanism for all data sources and consumers.

Automated infrastructure management for the cloud: PaaS support, including cluster configuration (TLS, load-balancing, autoscaling, etc.), start/stop controls, automatic installation of updates, and integrated monitoring.

GraphQL support: zero-code creation of GraphQL data APIs, with first-class performance by leveraging the Denodo query optimizer.

"Apache Zeppelin for Denodo" Notebook: Data scientists can construct narratives that combine queries, code, and text, to aid in data analysis and to help them to explain their work and share it with colleagues.

Redesigned data catalog with automatic recommendations and enhanced collaboration, profiling, vand search features

Business Benefits

HIGHLY ECONOMICAL

Integrate data reliably at a fraction of the time, cost, infrastructure investment, and rigidity of traditional data integration approaches such as extract, transform, and load (ETL) processes.

FASTER PATH TO VALUE

Deliver contextual and reliable information faster for actionable insights. Agile enterprises are more likely to be industry leaders.

LOCATION-AGNOSTIC

Deploy in any location – on-premises, cloud, or edge – without sacrificing scale or data governance capabilities.

BUSINESS-FRIENDLY

Abstract the complexity of modern data ecosystems (myriad sources, multiple formats, distributed, heterogeneous, diverse) from business. Expose data in the right format and using the naming conventions required by every type of user and application at almost no cost. Rapidly adjust to changes in requirements.

ENTERPRISE GRADE

Support multiple lines of business, multiple use cases, multiple persona, and thousands of users.

A DIGITAL MARKETPLACE

Enable a digital marketplace that empowers a community of analysts to find and use information assets quickly, which is essential in this age of self-serve analytics.

Denodo Platform 8.0 Capabilities Sheet

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
- Oracle (JDBC): 8i, 9i, 10g, 11g, 12c, 18c, 19c
- Oracle E-Business Suite (JDBC): 12
- PostgreSQL (JDBC): 8, 9, 10, 11

Sybase Adaptive Server Enterprise (JDBC): 12, 15
- MS Access (ODBC)

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

Parallel databases and appliances
- GreenPlum (JDBC): 4.2
- HP Vertica (JDBC): 7, 8
- Oracle Exadata (JDBC): X5-2
- ParAccel 8.0.2 (by using ParAccel 2.5.0.0 JDBC3g with SSL driver)
- Netezza (JDBC): 4.6, 5.0, 6.0, 7.0
- SybaseIQ (JDBC) 12.x, 15.x
- Teradata (JDBC): 12, 13, 14, 15
- Yellowbrick

Cloud Data Warehouse / RDBMS
- Amazon Redshift (JDBC)
- Amazon Athena (JDBC)
- Amazon Aurora (JDBC)
- Amazon DynamoDB
- Amazon RDS
- Azure Cosmos DB
- Azure SQL Database
- Azure Synapse Analytics
- Delta Lake
- Google Cloud SQL
- Google BigQuery (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
- Presto (JDBC)
- Databricks 2.x

NoSQL
- MongoDB
- Cassandra

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, 11

Data Lake Storage
- S3
- Azure Data Lake Storage
- Azure Data Lake Storage Gen 2
- Azure Blob Storage
- Google Cloud Storage
- Parquet (Distributed File System Connector)
- Avro

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

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 locally accessible or in remote filesystems, through FTP/SFTP/FTPS, and in clear, zipped and/or encrypted format.

Indexes and unstructured content
- CMS, file systems, pdf, word, text, email servers, knowledge bases, indexes
- Elastic Search 6.4, 6.7

Cloud, Saas,Web Sources with Simplified OAuth Security
- Amazon
- Google
- Google Sheets
- Facebook
- LinkedIn
- MS Azure Data Lake
- MS Sharepoint (by using the OData connector)
- MS Dynamics 365 Business Central / Customer Engagement
- Marketo
- ServiceNow
- Salesforce (SOQL)
- Twitter via APIs with simplified OAuth integration (1.0, 1.0a and 2.0)
- Workday

Active Directory as source or leveraging security
- LDAP v3
- Microsoft Active Directory 2003, 2008

MS Queues as data source and Delivery
- MQSeries
- SonicMQ
- ActiveMQ
- Tibco EMS

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

Denodo SDK for Custom Connectors
- CouchDB
- Lotus Domino

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

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

The following data sources have been successfully tested with Denodo using JDBC and ODBC drivers, WS/SOAP and WS/REST, and DenodoConnect adapters (not exhaustive list):
- Apache Solr
- Kafka Messages
- Hadoop HBase
- Hadoop HCatalog
- Hadoop HDFS
- IBM BigInsights
- Pivotal HAWQ

PUBLISHING OPTIONS

- SQL Based access via JDBC, ODBC and ADO.NET
- Web Services
  - SOAP
  - REST
  - OData
  - Open API (a.k.a Swagger)
  - GraphQL

- Oauth, OAuth 2.0 (JWT)
- SAML
- SSL
- WS-Security
- JMS listeners for message queues
- Denodo Scheduler for batch process and lite ETL

DATA CATALOG

- Web-UI for data discovery and exploration for business users
- Business Categories and Business Tags
- Full search capabilities on metadata and actual data
- Query wizards for customizing datasets
- Export to CSV, Excel and Tableau Data Extracts
- Collaboration features (save and share)
- Export to a shared sandbox for IT review before final publication for global use
- Automatic personalized recommendations and shortcuts to recommended datasets
- Endorsement of datasets, comments, warnings, etc.
- Usage statistics: who uses what data, when and how
- Profiling information
- Smart search with smart ranking of results

PERFORMANCE OPTIMIZATIONS

- Smart Query Acceleration for Analytics
  - Aggregation Awareness, Summaries
- Massive Parallel Processing (MPP) integration 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
- Multi-mode caching: full, partial, incremental or total refresh, event-based or scheduled, configured at the view level, incremental queries for SaaS sources

CACHE AND DATA MOVEMENT OPTIONS

- 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

MASSIVE PARALLEL PROCESSING 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 including lineage information to data governance tools

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
- Masking
- Custom policies for specific security constraints and integration with external policy servers

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
- Leverage existing source privileges

DATA MODELING

- Design Studio: Web-based development studio for data modeling (desktop version still available)
- Bottom-Up and Top-Down (through Interface Views)
- Denodo Model Bridge: Integration with third-party modeling tool
  - 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 (DMT) integrated in the Solution Manager, Denodo Query Monitor, Denodo Monitor Reports: all monitoring information is delivered via Denodo tools for real-time and historical analysis. Historical dashboards can be created using the monitor information
- 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)

OPERATIONS

- Solutions Manager to automate operations and promotions tasks
  - Centralized management and distribution of updates to clients
  - Centralized management of license keys
  - Defining 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.
  - Installation of Denodo updates on clusters with several servers, without downtime.
  - 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
  - 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
  - Auto-scaling support both in AWS and Azure

USER INTERFACES

- Central Web Console: integrated access to all Denodo UIs with SSO (Kerberos, SAML, OpenID and OAuth)
- Solution Manager Web-based UI: Centralized UI for administrators to manage deployments and promotions, including automatic management of cloud infrastructure (AWS)
- Design Studio: Web-based Development Studio, drag-and-drop and low-code developer studio

geared to data-oriented developers such as data engineers, power users, and citizen integrators; publish data services with a few clicks.
- Desktop Dev. Studio (VDP Admin tool)
- Data Catalog Web-based UI: Easy-to-use interface for business-oriented users such as data stewards, data analysts, and citizen analysts.
- Diagnostic and Monitoring Tool: centralized Web-based UI for monitoring, auditing and troubleshoot for data engineers and administrators.

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
- UNIX (64-bit platforms): Sun Solaris
- 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.