Denodo Platform 5.0

In today’s information economy, organizations rely heavily on the availability and quality of data at all levels to support investments in new business capabilities, business intelligence, compliance and for creating information-driven strategic advantage. This requires moving away from largely siloed and vertical-focused data management solutions into architectures that offer a unified data layer through virtualization for the agile delivery of data services across an increasing number of physical data sources and to multiple applications and users.

Denodo Platform 5.0 responds to the need of dealing with more data types, ubiquitous data and bigger data volumes, reducing data latencies, reducing data replication and delivering more agility and value than traditional technologies. The platform is an integrated Data Virtualization platform that creates, manages and delivers virtualized data services at both project and enterprise levels with speed, agility and high performance.

Denodo Platform 5.0 connects to data sources – be they structured, semi-structured or unstructured, internal or external – and combines them into logical / virtual data services to provide unified access and integrated delivery through a single “virtual” data layer and publishes these to consuming applications in real-time (right-time). It includes key capabilities for real-time query optimization supported by intelligent caching and scheduled data orchestration, unified data governance and quality, and ability to deliver data services in multiple formats with managed security and service-levels.

Denodo Platform 5.0 is an integrated software platform that delivers these fundamental capabilities required for world-class Data Virtualization:

- **Universal Data Access**: Read/write from any source or data type including Legacy, NoSQL and Cloud. Web Automation and Indexing automatically navigates and extracts data from Web and unstructured sources into structured views for integration.

- **Unified Virtual Data Layer**: Builds powerful transformations and relationships using an integrated modeling and execution environment to normalize, transform, improve quality and relate data across heterogeneous source types using common metadata and semantics. An extended relational data model allows disparate data types to be represented natively in the virtual layer minimizing effort and maximizing performance.

- **Universal Data Provisioning**: Expose the combined information as reusable Linked Data Services in multiple formats (SQL, SOAP WS, fully RESTful Linked Data, XML, JSON, XHTML etc.) and supporting hybrid delivery modes (virtual real-time, cache, batch, message-based physical movement) to consuming applications.

- **Agile High Performance**: Advanced real-time optimization supplemented by intelligent Caching and Scheduled Batch for flexible mixed workloads. Supports read/write access with enterprise class reliability and scalability – even for web and unstructured sources.

- **Unified Data Governance**: Enterprise-wide single entry-point for data and metadata management, security, audit, logging and monitoring enabled through built-in tools and instrumentation as well as integration to external data management tools.

- **Agile Development**: Rapid ways to deliver pervasive, self-service data services using graphical, wizard-driven UI and discovery tools. Hide complexity to application developers and business users; Decouple consuming applications and data sources; Allow easy creation, extension and use of data services.

**Benefits**

Denodo Platform 5.0 provides key benefits and advantages when compared to custom coding or traditional data integration solutions by delivering "more-for-less": more data access and provisioning in less time and cost.

- **All data leveraged**: Combine disparate data (Enterprise, Web, Big Data, NoSQL, Unstructured) to create meaningful new business data services.

- **Right-time integration**: Fresher information from the source of truth, balanced with performance.

- **Cost saving**: Allows physical replication, but only as needed. Saves HW, SW and people costs.

- **Faster time to market**: A library of high quality data services speeds development on new projects.

- **Data Services reuse**: Data is leveraged efficiently across transactional, analytical and informational applications to deliver the benefits of virtualization.

- **Unified ease of use**: A unified access platform to graphically model all data sources and publish and manage data services that is easy and intuitive.

- **Scalable performance**: Advanced automatic query optimization with transparent manual control, latency balancing, and modular clustered scalability.
Features

**DATA ACCESS:** Point-and-click adapters to Enterprise, Web / Semi-structured and Unstructured data in any format or location. High-performance optimized adapters for all the main sources. Bi-directional read/write.

- Relational databases such as Oracle, DB2, Sybase, MS SQL Server, MySQL, PostgreSQL, Informix, MSAccess.
- Parallel databases and appliances such as Teradata, Netezza, Oracle Exadata, Sybase IQ, ParAccel.
- Multidimensional OLAP engines such as SAP BW, MS SQL Server Analysis Services, Mondrian, Essbase.
- Enterprise Applications such as Oracle E-Business suite, SAP R3 / ECC, Siebel, Peoplesoft, Salesforce.
- Mainframe / Legacy connectivity including Adabas, IMS, DB2, TN5250 / TN3270. Plug-in architecture for use of third party mainframe / legacy adapters.
- SOAP / REST Web Services and data feeds, including support for XML, RSS, ATOM, JSON and CSV formats.
- Semantic repositories in Triple Stores / RDF accessed through SPARQL endpoints.
- Connect and introspect LDAP and Active Directory services as source data (apart from security access).
- Big Data / NoSQL databases such as Hadoop, Hive, HBase, MongoDB, CouchDB, Neo4J, MarkLogic.
- Cloud, SaaS sources including Salesforce, Google, Amazon, LinkedIn, Facebook, Twitter via APIs; Any Website, Form, WebApp via browser automation.
- Connector SDK to access any custom application through API and procedural interfaces.
- Sophisticated tools to expose Web, Semi & Unstructured data as virtual relational data/service.

**WEB, SEMI/UNSTRUCTURED DATA INTEGRATION:** Automates extraction and integration of less-structured data from Web sites, forms, applications, PDF, MS Word

- Flexible Automation of Web integration processes modeled using a library of pre-built templates and components for workflow, navigation, extraction, and structuring of Web and semi-structured data.
- Automatic Web Navigation: Handles dynamic sites with AJAX, JavaScript, authentication, secure servers, sessions, cookies, popup windows or sequences.
- Data Extraction: Example-based heuristic extraction of semi-structured content into defined schema from dynamic Web 2.0 content, PDF forms, documents.
- High Performance: Parallel Execution using MS IE or Denodo browser pools; Smart browsers load only optimal navigation sequences to minimize memory.
- Content Integration: CMS, file systems, SharePoint, Email servers, knowledgebase, index, ontologies.
- Search/Index: Built-in tools expose unstructured data as inverted indexes or histogram databases.

**DATA INTEGRATION:** Transform, clean, combine disparate information into virtual canonical business views.

- Extended Relational Model: Natively represents and seamlessly combines relational, hierarchical, NoSQL and semantic representations of data into abstracted relational model; Takes into account source capabilities and constraints for query optimization.
- Metadata-driven Integration: Graphical wizard-driven UI and tools to introspect siloed data, create unified logical data views, transform, cleanse, group, aggregate, modify output, define data workflow, etc.
- Top-Down and Bottom-up Modeling: Interface Views allows specifying contract-first schema for high level business entities, which later can be linked to bottom-up implementation views and become fully executable. Parallel modeling utilizing views created from sources, other virtual views, imported from enterprise modeling tools or industry data models.
- Linked Data / Associations: Establishes relationships between virtual data entities using referential constraints (Primary Keys, Foreign Keys, Varying levels of Multiplicity, Conditionals); Associations can be introspected from external metadata/model tools (ERWin, ERStudio, Rational …) for governance, browsed/traversed by users manually, or programmatically by applications using navigational queries.
- Transformation, Quality, Matching: Large library of built-in functions for SQL, XSLT, XQuery, XPath, Java, and Semantic transformations with ability to plug-in external tools or custom functions in every step of query execution lifecycle. Built-in data workflow tool for complex transformation / data quality processes.
- Semantic Integration: Transform, relate and merge unstructured data with structured using text mining, taxonomy filters, semantic tools like textual similarity.
- Data Modeling to shape output, flatten or create new hierarchical data structures to match target schema.
- Bidirectional integration supporting read, write and transactions (2-phase commit and XA transactions).

**DATA PROVISIONING:** Deliver virtualized data as SQL views, data services, portlets/widgets to suit every need.

- Bespoke, optimized SQL views of unified virtual data layer accessed via ODBC, JDBC and ADOM.NET.
- Linked Data Services: Fully RESTful compliant interface exposes enterprise data assets as a unique URI accessible via search, browse or query drill-down using standard Web protocols, interfaces, and HTTP verbs (GET, POST, PUT, DELETE); Output in XHTML, XML, JSON for human and/or machine consumption.
- Data Services for SOA: Publish SOAP web services that conform to contract-first schema; Support XSLT tools, WS-security, ESB/JMS access, SOA catalog.
- Publish SharePoint WebParts, Java portlets, AJAX widgets, RSS to use in major portal/mashup servers.
- Publish and subscribe to data via JMS (with JSON support), including MQ- Series, Sonic MQ, Active MQ.
Deliver data using semantic formats: ability to answer SPARQL queries returning RDF (via D2R mappings).

**SCHEDULER & DATA ORCHESTRATION:** Enable complex, hybrid integration processes, integrating ETL within a broader Data Virtualization approach.

- Delivers reliable, high performance virtual data services through balance orchestration of real-time, cached, scheduled batch or hybrid execution modes.
- Materialize unified data views by exporting to databases, warehouses, flat files, Excel, XML, etc. using built-in Denodo Scheduler or external ETL tools.
- Support persistent tasks through continuation of query after restart, transparent retries in case of failures, intermittent Web access or human intervention.
- Task dependency, modeled via data orchestration workflow, allows linked tasks to start only when others have appropriately finished.

**DATA GOVERNANCE:** Data and Data Model assurance to ensure consistent, meaningful data services to users.

- Metadata Repository with multiple visualization (tree view, linked data, attribute origin, source impact, catalog search, etc.); Includes metadata API, model export and introspection from external systems.
- Discover, introspect, and transform source metadata. Refresh or propagate source metadata when it changes. Flatten or create new hierarchical data structures.
- Contract-first, top-down and bottom-up modeling for greater flexibility, parallel work, change governance.
- Data Lineage, Change Impact, Dependency analysis, metadata migration tools, version control, granular policy-based tiered security deliver a controlled data virtualization and enterprise data services capability.

**DEVELOPMENT:** Easy to use, enterprise-class tools.

- Graphical, wizard-driven UI for all functions and modules. Also, documented scripting for advanced users.
- Platform extensions enabled via Eclipse IDE plug-in for developing, testing, debugging and deploying custom functions, connectors and stored procedures.
- Integration with Version Control Systems (checkout, commit, update of virtual entities) with automatic dependency control directly from within Denodo.
- Graphical support for lifecycle process management (development, staging, production) or geographically dispersed environments.

**PERFORMANCE, SCALABILITY, RELIABILITY:** Integrated query optimization with cache and ETL support for agile high performance. Flexibility to scale in steps to enterprise class needs. Workflow management.

- Query Optimization: Intelligent and automated techniques (cost & rule-based using information from source introspection, query capabilities & constraints) including asynchronous delivery, query delegation, automatic parallel access to data sources, support for high-performance sub-queries, etc.
- Tunable optimization: Visual trace to inspect all details of query, before, during and after execution by view and by source with manual plan override of automatic strategy selections for optimized query.
- Advanced Multi-mode Cache: Configurable view-by-view with full, partial or incremental loads, with Intelligent Query Matching to answer queries directly from cache. Supports push-down of complex queries to cache for high-performance. Cache refresh can be based on schedule, event-trigger or expiration-time. Disk and in-memory databases supported for cache.
- Built-in Scheduler and external ETL support for data prefetch to cache or materialized source improves query delegation and balances source latencies.
- Database indexes and primary keys are introspected from sources and exposed to consuming applications and reporting tools for high-performance. Ability to expose these in virtual and cached views also.
- Web Services and Cloud optimizations include automatic unfolding of single queries into multiple parallel accesses, advanced session-reuse mechanisms, etc.
- High Availability: Support clustering, load balancers and federated deployments to distribute server workload and share metadata across virtualization layer.
- Robustness: Automated swapping, support for XA-compliant endpoints, fully transactional catalog storage, custom policies for workload management.

**SECURITY & MANAGEMENT:** Secure and differentiated role-based access control to data services and sources; Enterprise tools for management, monitoring dashboards

- Role-based authentication and authorization using LDAP, Active Directory and/or built-in user directory
- Security & Workload management using custom policies to restrict or constrain service-levels and workload based on external factors (source load, network, time); Integrates with external systems for security, access control, SLA policy management.
- Support for fine-grained security at virtual data view, column, or row level and for masking sensitive data.
- Also supports pass-through authentication to leverage security infrastructure in the datasources.
- Support for importing and exporting encrypted data; Communication between modules can be encrypted and authenticated using SSL also.
- Firewall support: All components can be distributed in different network segments.
- Denodo Dashboard with multiple tools for lifecycle management of servers, clusters and HA configurations. Metadata migration tools. Denodo Monitor to view real-time queries and historic audit/logs; Support JMX, SNMP, WS-Management standards to integrate with leading external monitoring tools including HP Openview, IBM Tivoli, Microsoft RM, Nagios, etc.
DENODO PLATFORM REQUIREMENTS

- Denodo Platform is a complete data virtualization and enterprise data services solution that runs as a stand-alone server and includes several components: design tools, connectors, virtualization server, runtime components, Eclipse IDE, administration tools, monitoring dashboard, etc. Denodo Platform Control Center provides a single-point of management of the various components.

- Operating System: Any OS with a Java 6 Virtual Machine on both 32 and 64-bit architectures. Includes most versions of Windows including XP, Vista, Windows 7, Windows 2003, Windows 2008. Most versions of Linux, including RedHat, CentOs, Fedora and Ubuntu. Most versions of UNIX. Mainframes OS including zOS.

- Directory Services: LDAP v3; Microsoft Active Directory 2003, 2008; Denodo embedded service.

- Browsers: Internet Explorer 8.x, 9.x; Firefox 3.0.x, 3.1.x, 3.5.x; Denodo Browser.

DEPLOYMENT PATTERNS

- Denodo Platform can be deployed in the data center or private cloud (either in physical or virtual servers) or in the public cloud (e.g. Amazon EC2). Strong Web Services support makes Denodo very cloud-friendly.

- Multiple Configurations supported including: Basic single server configuration; Basic server with proxy configuration for firewalls; High Availability Clusters with load balancer in either active-passive (hot standby) or active-active (horizontal scaling) configurations; Clusters with shared cache or distributed local cache; Geographically distributed servers environments; Multiple Denodo instances in peer-to-peer or multi-layered environments.

- Denodo Data Virtualization can also be embedded in other product architectures (for example Reporting & Dashboards, Single-View applications, Information Services delivery platforms, BPM/Workflow).

Denodo In Your Enterprise

Most organizations have siloed data sources that multiple users and applications are trying to access. This has created multiple point-to-point integrations and the acquisition of multiple integration middleware - ETL for datawarehousing and business intelligence, BPM/ESB for process integration, portals and file sharing systems for collaboration and teamwork, and search and indexing for unstructured content. Below is a high-level view of how Data Virtualization fits in your enterprise information strategies and the tactical and strategic roles it plays.

At the tactical level, Denodo adds to your integration toolkit the capability for agile, real-time data integration.

At a strategic level, Denodo enables a truly Unified Data Layer that abstracts and virtualizes all enterprise data assets into canonical business views that can be exposed as Linked Data Services or bespoke SQL. In this way it integrates existing middleware and data tools (modeling, data quality, etc.) to create a high performance and agile information architecture to serve all of your applications.