Expert Trail: Modeling and Metadata Organization

Revision 20211109

CONTENTS

1 LOOKOUT

Expert trails guide Denodo users through all the relevant materials related to a specific topic, including official doc, KB articles, training, Professional Services offering, and more. The main goal is to give users a single place with references to all the information that they need to become a Denodo expert on any specific topic.

When starting with a data virtualization project, an importing topic is to know how to do the modeling and how to organize the metadata. This Expert Trails guides you through those topics and provides best practices to become an expert.

2 THE HIKE

Stage 1: Define the naming convention and metadata organization

The first step of this stage is to define the naming convention that should be used in the data virtualization project. The Knowledge Base article VDP Naming Conventions provides an overview of the different layers: Connectivity, Integration, Business Entities, Report Views and Data Services and a recommended folder structure reflecting those layers.

After having an understanding of the different layers and naming conventions, it should be defined how to organize the projects within the virtual databases.

If there are common data sources and views shared by different projects they can be organized separately either in a common database or in a common folder, so that they do not need to be recreated several times, but only once and other projects can make use of them.

For organizing the projects themselves, there are two main approaches available: the folder-oriented model vs the database-oriented model. The presentation Denodo Development Best Practices gives an idea on the pros and cons of each approach.

Stage 2: Top-down vs bottom-up modeling

Once it is decided on how to organize the data virtualization project, it is important to be aware of the different concepts used for modeling the views themselves. In the Bottom-Up approach it is possible to start from the "bottom" by creating the data sources and base views and grow from there. The final views used by consumers or business users might not be defined at this point.
In the Top-Down modeling, however, it is possible to create interface views which allows to define a view schema for contract-based development. Once the interface view is created with a specific view schema, the implementation of this interface will then be later taken care of by the development team. It means that different teams can agree on a specific view schema on the final views, even though the base views and integration views needed for that view are not created yet.

Of course, it is possible to use both approaches in combination.

The Denodo Model Bridge allows to integrate models from third-party modeling tools (such as ER/Studio Data Architect, ERwin Data Modeler and others) into the Virtual DataPort Server by creating the interface views and associations related to these existing models.

Stage 3: Build the Connectivity

At this point, the project structure and naming conventions are defined and there might already be some interface views imported via the Denodo Model Bridge. The next step is to proceed by creating elements belonging to the first layer: Connectivity. These are the data sources and base views.

The Expert Trail: DataSource Connectivity is useful to know more about what kind of data

sources can be integrated with Denodo and how to do it.

The second part of the Connectivity layer, after having created the data sources, is to create base views. Depending on the data sources that should be integrated, the steps are slightly different. The documentation section Creating Data Sources and Base Views explains how to create the base views for the different types of data sources.

Stage 4: Combine the data

After having created the first base views, the next step is to combine and transform them to new views, which are called derived views or integration views. These views build the foundation for the upper and final views (business views, report views, data services). Speaking in terms of the project layers introduced in Stage 1, they can belong to the integration but also to the business entities layer.

While designing the model, there should be a compromise between reusability and maintainability. If for each transformation step (join, group by, selection, flatten...) a separate integration view is created, chances will increase that those views can be reused for multiple final views, however, the maintenance of those views will be more time consuming. The opposite would be to bundle many transformation steps into a single view, which makes the maintenance easier as it will result in less views, but decreases the chances that this view can be reused for other final views as well. The recommendation is to find a middle way between those two approaches. The presentation Denodo Development Best Practices also explains details about how to create the model.

The model definition can affect the performance of the queries for the derived views. For more information on this check the Expert Trail: Query Performance Optimization.

Stage 5: Consume the data

Now that the model is built, the next question is how to access it. The Denodo Platform provides different ways to consume the data from third party clients. The main connectivity methods are through JDBC, ODBC and the RESTful Architecture, which consists of the GraphQL Service, the OData 4.0 service and the RESTful Web Service.

In addition to that, it is possible to discover and visualize the model by using the Data Catalog. The Data Catalog is a component intended mainly for business users in order to discover different data resources in the organization through browsing or searching. When clicking on specific views, it is possible to see the data lineage and query the views from there if the user has the necessary privileges.

3 EXPLORATION

Fill up your backpack with additional gear:

Modeling

Official Documentation	●	Denodo Model Bridge - User Manual
KB Articles	●	Best Practices to Maximize Performance I: Modeling Big Data and Analytic Use Cases
	●	Using the Create Base View From Query Option
	●	How to model slowly changing dimensions in Denodo
	●	VDP Naming Conventions
Webinars	●	Denodo Design Studio: Modeling and Creation of Data Services

Metadata Organization

KB Articles	●	Groups of Elements in Denodo
	●	Denodo Development Best Practices | Recommended Project Structure
	●	Denodo Development Best Practices | Virtual Databases Organization

Associations

Official Documentation	●	Why You Should Define Associations Between Views
	●	Creating an Association
	●	Associations of Views in the Data Catalog
	●	Navigating Associations in the OData 4.0 Service
KB Articles	●	Associations in Denodo

Interfaces

Official Documentation	●	Creating Interface Views
Additional Resources	●	Defining contracts/interfaces - Tutorial | Denodo Community Site

Data Catalog / Data Discovery

Official Documentation	● [Introduction to the Data Catalog](#)
Webinars	● [Data Catalog for Better Data Discovery and Governance](#) ● [AI/ML: Automating Data Discovery with AI/ML-Enabled Data Catalog](#) ● [How to Simplify Data Discovery with Machine Learning Data Catalog? (I)](#) ● [How to Simplify Data Discovery with Machine Learning Data Catalog? (II)](#) ● [How to Simplify Data Discovery with Machine Learning Data Catalog? (III)](#)
Additional Resources	● [Data Discovery - Tutorial	Denodo Community Site](#)

4 GUIDED ROUTES

4.1 DENODO TRAINING COURSES

Denodo training courses provide expert data virtualization training for data professionals, including administrators, architects, and developers.
If you are interested in "Modeling and Metadata Organization" Expert Trail you should enroll the following course/s:

● **Data Modeling with Denodo**: This course provides concepts of using Denodo for Data modeling, which is the method of defining and analyzing the data requirements needed to support the business functions of an enterprise.

4.2 TECHNICAL ADVISORY SESSIONS

Denodo Customers with active subscriptions have access to request Meet a Technical Advisory sessions.
These are the sessions available related to Modeling and Metadata Organization.

Development Methodology	**Modeling: Metadata Organization**	Recommendations on defining: - How to organize elements - Database Organization Approach - Naming conventions
	View Modeling: Best Practices	Recommendations on defining: - Business model: Define Canonical Model, Report presentation layer, Relationships - Integration with database modeling tools (i.e. Embarcadero, Erwin, etc.) - Top-Down and Bottom-up development approaches
	Data Delivery: Architecture Overview & Best Practices	- Define policies for accessing the Denodo Platform through the different available interfaces. - Data services best practices: RESTful architecture, OData, GraphQL, Swagger, etc.
Data Catalog & Governance	**Data Catalog: Overview & Best Practices**	Capabilities review and general best practices that can be followed to kick start your cataloging activities: - Metadata documentation and classification (tags, categories) for data stewardship - Advanced metadata and content search - Data exploration: Web-based query wizard for non-technical users, personal reports, and sharing options - Usage statistics

4.3 PROFESSIONAL SERVICES

Denodo Professional Services can help you at the start or any part of your query performance trail. You can find information about the Denodo Professional Services offering in:

[Professional Services for Data Virtualization | Denodo](#)

- Development Quick Start
- Implementation Assistance

If you are a Denodo customer, you can reach out to your Customer Success Manager for details about any Guided Route that you need.

5 BIG HIKE PREP CHECK

Let's see if you are ready to start your big trail. Take this 4-question questionnaire to check your readiness for an enjoyable hike.

Read the questions below, think about the solution and check if you got them right by looking at the solution. Have you become an expert?

1. The Data Catalog can be used to browse through the views and discover data, mainly intended for business users. Does it also allow you to customize queries and export data from there?
   Click here to check if you got it right

   > Yes. Customizing queries is possible through the tab "Query" when selecting a view from the Data Catalog. For example, you can add expressions and aggregations to the query and save them for later use. Furthermore, it is possible to export the data from there to different formats such as HTML, CSV and Excel.

2. What are the benefits of having associations defined?
   Click here to check if you got it right

   > There are several benefits of having associations defined in Virtual DataPort as seen in the documentation section Why You Should Define Associations Between Views:
   > - Performance: Some optimizations can only be applied if the associations are defined.
   > - For external clients: The associations are exposed to external clients and that information can be used by the client applications to execute more efficient queries or to suggest views to be added as a join and automatically create the join conditions based on the association.
   > - Metadata/Model: The associations provide additional information to your semantic model that could be helpful when accessing it through an external governance tool.
   > - Data Catalog: The Data Catalog allows business users to discover related views through traversing the associations or through the "Relationships" tab.
   > - RESTful Web Service: The RESTful Web Service allows you to browse through views that have an association defined between them.

3. Developers manage a lot of elements in a Data Virtualization project. Can you organize the following elements in the correct order?
   Connectivity
   Data Services
   Report Views
   Integration
   Business Entities
   Click here to check if you got it right

   > The correct order is
   >
   > Connectivity - (Data sources and Base views)
   > Integration - (Combination and transformation aka derived views)
   > Business Entities - Canonical model
   > Report Views - Pre-built reports
   > Data services - Finally the web services

> This project structure is described in the presentation Denodo Development Best Practices under the section "Recommended Project Structure".

4. Is there any other usage of the interface views other than for Top-Down modeling?
Click here to check if you got it right

> Yes. In addition to the Top-Down modeling the interface views can be used in the scenario of a data source migration. Let's suppose that data stored in Oracle on-prem is planned to be migrated to the cloud. An interface view can be used that implements the current view (coming from Oracle) and, at a later step, it can be easily switched to the new cloud version of the data. Once the implementation is changed, end users will not notice any difference as they continue to query the same integration view, i.e. the integration view decouples the consumers from the actual data sources.