Expert Trail: DataSource Connectivity

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.

"How to connect to my Data source from Denodo?" is usually the first question that comes to our mind when starting a project with Denodo Platform.
In this Expert Trail, you will learn different types of options to access your data source, configuration techniques, and some of the advanced features in data source configuration.

2 THE HIKE

Stage 1: Different ways to access data in real-time

The first step in a Denodo project is to create data sources to connect to various data repositories, API, etc. Other steps on the Denodo project are described in the Expert Trail: Development Lifecycle and how the metadata organization should be set up is described in the Expert Trail: Modeling and Metadata Organization.

The following are the different types of data sources available in Denodo:

- **JDBC**: to access relational databases like Oracle, Snowflake, etc.

- **ODBC**: to access ODBC sources
- **Web Service**: to access the SOAP Web service
- **Multidimensional DB**: to obtain data from multidimensional databases like Azure Analysis Services, Oracle Essbase, etc.
- **XML**: to access the XML files
- **JSON**: to access data from the JSON files
- **Delimited File**: to access data from delimited text files like CSV
- **Excel**: to access data from Excel files
- **LDAP**: to delegate authentication tasks to an LDAP server
- **Salesforce**: to access data from Salesforce REST API
- **BAPI**: to obtain data stored in SAP ERP and other SAP applications.
- **Custom**: to access ad-hoc wrappers developed in Java for which no standard connector is available.

The document Creating Data Sources and Base Views explains how to create a data source for each type mentioned above.

There are few predefined custom connectors to access some of the data sources like MongoDB, file formats stored in HDFS, S3, Azure Data Lake, etc, that can be downloaded from the DenodoConnects section of the Support Site.

Stage 2: What do you need to configure the data sources

Let's focus on different configuration parameters required to create data sources in Denodo.

- For the **JDBC data source**, you need a JDBC driver, driver class, connection URI. The Supported JDBC Data Sources documentation provides the details of the JDBC adapters and the JDBC drivers included in the Denodo installation. You could use the option File > Extension management from the Administration Tool to upload a custom JDBC driver or to upload a JDBC driver not included by default in the Denodo installation. The section Importing a JDBC Driver explains how to do this.

- For **ODBC Sources** the most common connection type is the use of DSN. You could use the existing DSN of the system where the Virtual DataPort server is installed.

- When you create a data source of the type **XML, JSON, Excel**, **Direct connection type of ODBC source, or delimited text file (DF)**, you have to specify a path to the data files. The document Path Types in Virtual DataPort explain in detail the different path types available in Virtual DataPort.

- For the **SOAP web service**, the WSDL file that contains the Web service specification needs to be specified.

- For **SAP** (SAP BW 3.x (BAPI) or SAP BI 7.x (BAPI)) and **BAPI data source**, you need to install the SAP Java Connector and Grant Privileges in SAP for BAPI Data Sources and SAP BW for Multidimensional Data Sources.

- For **Mondrian, Microsoft SQL Server Analysis, and Azure Analysis Services** the XMLA URI endpoint is required so the XMLA interface has to be enabled.

- For **Salesforce**, the main aspect to configure is the authentication flow. This connector uses the protocol OAuth 2.0 and supports this authentication flows:
    - Web server flow
    - Username and password flow

In the case of the custom data source, you would need to develop a custom wrapper using the APIs provided by Denodo. The section Developing Custom Wrappers explains how to develop them.

Stage 3: Authentication

Now, let's focus on different authentication methods available in the data source configuration

- **Use login and password**: This is the easiest way to connect to the data source. This option is available in data sources such as JDBC, ODBC, BAPI, Multidimensional DB, and LDAP.
- **Use Pass-through session credentials**: If selected, Virtual DataPort will connect to this source using the credentials of the user that executes the query. This option is available almost in all data sources except LDAP and Salesforce data sources. Please note that using this option is depending on the capabilities of the source underneath. To utilize this feature provided by Denodo, the underlying source needs to support it. The section Considerations When Configuring Data Sources with Pass-Through Credentials explains in detail the consideration using the Pass-through.
- **Use Kerberos**. This network authentication protocol is an option available in JDBC, BAPI, and Multidimensional data sources. In case of using a JDBC data source, read the section Connecting to a JDBC Source with Kerberos Authentication for more information about this.
- In the case of using **HTTP connections** for the data source such as XML, JSON, Excel, and delimited text file (DF) data source, the supported authentication methods are as follows: Basic, Digest, Mutual (two-way SSL), NTLM, SPNEGO (Kerberos), OAuth 1.0a and OAuth 2.0. The documentation Authentication in HTTP Paths explains in detail.
- The **Salesforce** data source uses the OAuth 2.0 authentication protocol to securely access data.

More information about Security can be found in the Expert Trail: Security.

Stage 4: Advanced features in different data sources

Next, we will see some of the advanced configurations to optimize the data source.

Connection   **pooling:**

To minimize the cost of opening connections, Denodo utilizes an optimization technique called connection pooling. A connection pool is a cache of database connections maintained so that the connections can be reused when future requests to the database are required. This option is available in data sources such as JDBC, ODBC, LDAP, Web services, Multidimensional, and BAPI data sources. The Pool of Connections of the JDBC Data Sources documentation explains more in detail.

Source Configuration:

The Source Configuration in the data source specifies certain characteristics of the underlying sources, such as their support operations. For instance, if you have created a custom delegatable function then you can append this function to the Delegate scalar functions list property to delegate the function to the data source. This option is available in JDBC and ODBC data sources. The Data Source Configuration Properties documentation explains in detail each property available in the data sources.

Driver Properties:

The Driver properties are the list of name/value pairs that will be passed to the JDBC driver when creating a connection with the underlying database. For instance,

'v$session.osuser'='@{USER_NAME}' in the Oracle driver properties will assign the Virtual Dataport session username to oracle JDBC property v$session.osuser. In the case of JDBC, the option "Driver properties" will allow you to set the driver properties. Similarly, for the ODBC source, click "DSN Properties Configuration" to add properties to the connection URI.

Bulk load API:

JDBC data sources can be configured to use the bulk load APIs and this option useful for writing data into a database. This option is useful for caching, remote tables, summaries, and data movement optimization. The Bulk Data Load documentation explains more in detail.

Proxy:

For the data source that establishes an HTTP connection if the HTTP requests are sent through a proxy, Virtual DataPort provides an option to configure the proxy server information. The documentation Default Configuration of HTTP Proxy explains in detail how to configure these default values.

Secure Network Communications (SNC):

Secure Network Communications (SNC) provides stronger authentication and encryption mechanisms than the default security options of SAP. This option is available in Multidimensional and BAPI data sources for SAP.

Check Certification:

This option is available in the data sources that establish HTTP connections such as XML, JSON, Excel, delimited text file (DF) data source, Web Service datasource. You could utilize this option if the service requires SSL client authentication. The Importing the Certificates of Data Sources (SSL Connections) documentation explains how to do this.

Filter:

Virtual DataPort can access data files that are compressed or encrypted. This option is available in XML, JSON, Excel, and delimited text file (DF) data sources. The Compressed or Encrypted Data Sources documentation explains in detail this option.

Pagination:

This option will allow you to limit the number of records it returns per request. You can configure DF, JSON, and XML data sources to retrieve data from a REST API paginated. The section Pagination of Virtual DataPort Administration Guide elaborates in detail this option.

Stage 5: Error analysis

In the final stage, we will focus on how to debug the errors that may occur during the data source connection.

Let's take an example: Following is the error you may receive when trying to make a connection to SQL Server " *IOException: Connection timed out error*".

First, look into the vdp.log file located under <DENODO_HOME>/logs/vdp directory for the detailed error message. The error message suggests the data source is not reachable from the Denodo host machine. For further analysis, make sure there are no firewall restrictions or network issues. You can do so by doing a PING to the data source host or connecting from a different JDBC client. Similarly, if the issue is for a REST API try connecting to the same data source from POSTMAN to rule out connectivity issues.

Logging System:

Denodo Platform uses the library Apache Log4j 2 to log its activity. To debug the error and to analyze its root cause, you could look for generated logs. The Log column details Knowledge Base article describes the content of the VDP logs and Denodo Monitor logs that could provide insights into which log file what kind of information to debug.

Modifying the log level:

To get additional information from the log file, you could define different log levels for different log categories. The section Configuring the Logging System of Virtual DataPort Administration Guide and Using logcontroller and useful log categories provide insights on configuring the different log levels and some useful log categories.

More information about Monitoring can be found in the Expert Trail: Monitoring.

3   EXPLORATION

Fill up your backpack with additional gear:

JDBC

Official Documentation	● JDBC Sources
KB Articles	● Best Practices for Migrating SQL queries to Denodo VQL ● Denodo JDBC Connection Pooling Implementation ● How to configure the number of idle connections in a connection pool ● Scalar functions list in a data source ● Testing JDBC connections
Additional Resources	● JDBC - Drivers	Denodo Community Site ● Connecting to Data source - Tutorials	Denodo Community Site ● Connecting to Apache hive - Tutorials	Denodo Community Site ● Connecting to other Big Data systems - Tutorials	Denodo Community Site

ODBC

Official Documentation	● ODBC Sources
KB Articles	● Connecting Denodo to Azure Cosmos DB ● Closed connections from database cause error
Additional Resources	● ODBC - Drivers	Denodo Community Site

File type sources

Official Documentation	● XML ● Delimited file source ● JSON ● Excel datasource
KB Articles	● Using the From Variable Data route

	● Importing fixed-width files - Data Sources
Additional Resources	● Twitter, sentiment analysis and Denodo
	● Excel marketing file - Tutorials | Denodo Community Site
	● Joining Views from Disparate Sources - Videos | Denodo Community Site

LDAP sources

Official Documentation	● LDAP sources
KB Articles	● Connect to LDAP data source using SSL
	● How to configure a VDP database with LDAP authentication
	● LDAP authentication best practices
Additional Resources	● Creating LDAP data sources - Videos | Denodo Community Site

Web services

Official Documentation	● SOAP Web Service Sources
KB Articles	● Denodo query delegation and OBL fields
	● Using like operator in SOAP web services
	● Web service data source not returning any results when streaming output
Additional Resources	● Import a SOAP Web Service - Videos | Denodo Community Site

BAPI

Official Documentation	● BAPI sources
KB Articles	● Denodo and SAP Best Practices - Data Sources
	● Creating an SAP extension for the Denodo RFCReadTable Custom Wrapper
	● SAP, Accessing DSO tables Through BAPI

| Webinars | ● SAP Analytics Cloud: Do you already have live access to all data sources? |

Multidimensional

| Official Documentation | ● Multidimensional source |
| KB Articles | ● Denodo and SAP Best Practices
● SSAS as data source |

Salesforce data source

| Official Documentation | ● Salesforce sources |
| KB Articles | ● How to connect to Salesforce from Denodo
● How to use Salesforce Connect to access Denodo |

Custom data source

Official Documentation	● Custom sources
KB Articles	● How to integrate Amazon S3 with the Denodo Platform ● Accessing SAP SuccessFactors from Denodo ● How to connect to Microsoft Dynamics 365 from Denodo ● Microsoft SharePoint as data source ● How to connect to MongoDB from Denodo
Additional Resources	● Cloud sources - Videos | 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 DataSource Connectivity you should enroll in the following course:

- **Connecting and Combining Data with Denodo:** This course provides data developers with Denodo Platform 8.0 concepts, terminology, and skills needed to develop a data virtualization project.

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 Datasource connectivity.

Platform Administration	Data Source Integration	Assist in how to integrate with an out-of-the-box data source, i.e.: - Traditional relational databases (Oracle, SQLServer, DB2, MySQL, Postgres, etc.). - Data warehouse databases (Netezza, Teradata, Vertica, Greenplum, etc.). - In-memory databases (HANA, Oracle In-Memory, etc.). - Cloud databases (Redshift, Snowflake, AWS RDS, etc.). - SQL-on-Hadoop (Hive, SparkSQL, Impala, Presto, etc.). - Delimited files, logs. - Excel. - Web APIs and SaaS (SFDC, Google APIs, Workday, etc.). - SAP ECC. - Multidimensional cubes (SAP BW, Essbase, MS SSAS, etc.). - Hadoop HDFS files, Amazon S3 files. - NoSQL databases (MongoDB, Cassandra, Neo4j, etc.). - LDAP. - Semi-structured (Word, PDFs, websites, etc.)

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

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 trial. Take this 5-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. You have created a JSON data source to read a set of log files with a local path that includes date range, and the data source will read all the log files in order. If one of the dates mentioned in the specific date range is missing, the query will fail. Would it be possible to avoid this kind of error in Denodo?
Click here to check if you got it right

Yes, you could select the checkbox "Ignore route errors" of data source configuration to ignore the errors that occurred when accessing the file(s) to which the data source points. The section JSON Sources of Virtual DataPort Administration Guide explains in detail this option in the data source configuration.

2. What can you do when Denodo does not provide a standard connector to connect to your data source?
Click here to check if you got it right

Denodo provides connectors for the most commonly used interfaces, so review all the accesses to your data source to make sure that none of them can be used. However, if none of the out-of-the-box connectors and the DenodoConnects are useful for you, develop a custom wrapper and create a custom data source. The documents Creating Data Sources and Base Views and Denodo Connect Manuals provide detailed information about this.

3. Would it be possible to get the data source access information from the log? If yes, what was the exact log category that would be beneficial to retrieve information about XML data sources?
Click here to check if you got it right

Yes, the log category **com.denodo.vdb.engine.wrapper.raw.xml** would be helpful to get its data source access information in the logs. The Knowledge base article Using logcontroller and useful log categories provides information on the list of useful log categories.

4. You have created a JDBC data source to retrieve data from Oracle, and you want this data source to list its synonyms. Would it be possible to introspect its synonyms in Denodo?
Click here to check if you got it right

Yes, you could make use of the "Driver Properties" feature and add the driver property **includeSynonyms** with the value **true**. The section Database Specific Information provides detailed information regarding this.

5. How could you access a file when it is encrypted with the algorithm PBE with HMAC-SHA256-512 and AES-256?
Click here to check if you got it right

You could use the "Filters" option available in the data source configuration to decrypt the encrypted file. The section Compressed or Encrypted Data Sources explains in detail the different options that Virtual DataPort provides to access encrypted data files.