denodo

Profile
The case study is about a global investments bank dedicated to helping clients manage and service their financial assets throughout the investment lifecycle. The bank itself is composed of several lines of business such as capital markets, financing and liquidity, securities services, treasury services, investment management, wealth management, and data and analytics. As of March 31, 2021, the bank had more than $40 trillion in assets under administration, and more than $2 trillion in assets under management.

"

"The data virtualization layer built on the Denodo Platform has made data consumption fairly easy and straightforward at the bank while ensuring confidentiality and security of data through support for RBAC policies"

**- Director of Architecture
  and Data**

Democratizing Data to Enable a Data-Driven Culture at a Major US Investment Bank

To promote a data-driven culture and make it effortless for business users to leverage real-time data, the bank built a new data infrastructure referred to internally as the Enterprise Data Platform (EDP). The Denodo Platform is a key component of the EDP, as it facilitates the processing of large amounts of data per day and automates end-to-end data ingestion, data transformation, and data governance platform, providing full support for self-service analytics.

Business Need

With the EDP project, the bank aimed to meet the massive data and analytics needs of the organization by improving business processes and performance. "Our goal," said the director of architecture and data at the bank, "was to cast as wide a net as possible to aggregate enterprise data, so as to include flat files, real-time data, streaming data, regulatory reporting data, and batch swipes, and from that build a single source of truth." One of bank's key requirements for the EDP was fast onboarding, which meant that the platform needed to offer self-service capabilities, be fully automated, and enable the right tools and users to interact with the platform to determine what is happening to their data and understand how to better manage their data. Being a colossal organization, the bank also required the ability to serve multiple data stores and be able to process large volumes of data ranging from 800 million to 1 billion data points a day. The director adds, "We also wanted to have a fully automated mechanism to support data retention and archival activities with the ability to automatically move data from high-cost stores to low-cost stores." The bank also needed to be able to easily leverage descriptive metadata so that the EDP can make intelligent decisions about the data in real-time, to enhance the self-service user experience.

The bank implemented the EDP as an automated, end-to-end data ingestion, data transformation, and governance platform. The platform is required to support both batches as well as real-time streaming data sets. All the data is stored in a raw data zone, in an appropriate landing location depending upon the data type. The raw data sometimes needed to be enriched with reference data as soon as it was received, for which the EDP uses tools such as Pentaho and Spark.

The Solution

To ensure that the EDP met all of bank's requirements, the bank leveraged the Denodo Platform to establish an enterprise-wide data virtualization layer, above the company's disparate data sources, which acts as a single source of truth. The data virtualization layer aggregates data from the different source systems and makes it available to consuming applications without any physical data replication. The data modeling necessary for serving the visualization and reporting tools is also done in the data virtualization layer.

Since the bank is also interested in knowing how data moves from one application to another across the firm, tracking data lineage through the data virtualization layer is also important. The director says, "Every data point or message that comes into our system is tagged with bookkeeping metadata that enables the bank to essentially track every data point, measure internal metrics like latency and throughput for a particular data store, to keep up with SLAs." The bank also provisioned the role-based access control (RBAC) policies on the data virtualization layer, to manage authentication and secure the banks'

highly confidential data. The bank's RBAC API enables users to only see or use the data, through the data virtualization layer, for which they have the necessary credentials. If users wish to access a certain data set that they are not authorized to access, they must request access through the discovery function. In some cases, especially for very complex queries or where large volumes of data are involved, the EDP can bypass the data virtualization layer and go directly to the underlying store.

Figure1: The Architectural layout of the EDP, with the Denodo Platform providing a data virtualization layer to be consumed by visualization and reporting tools. The Denodo Platform is also a key component of metadata, governance, and data catalog services.
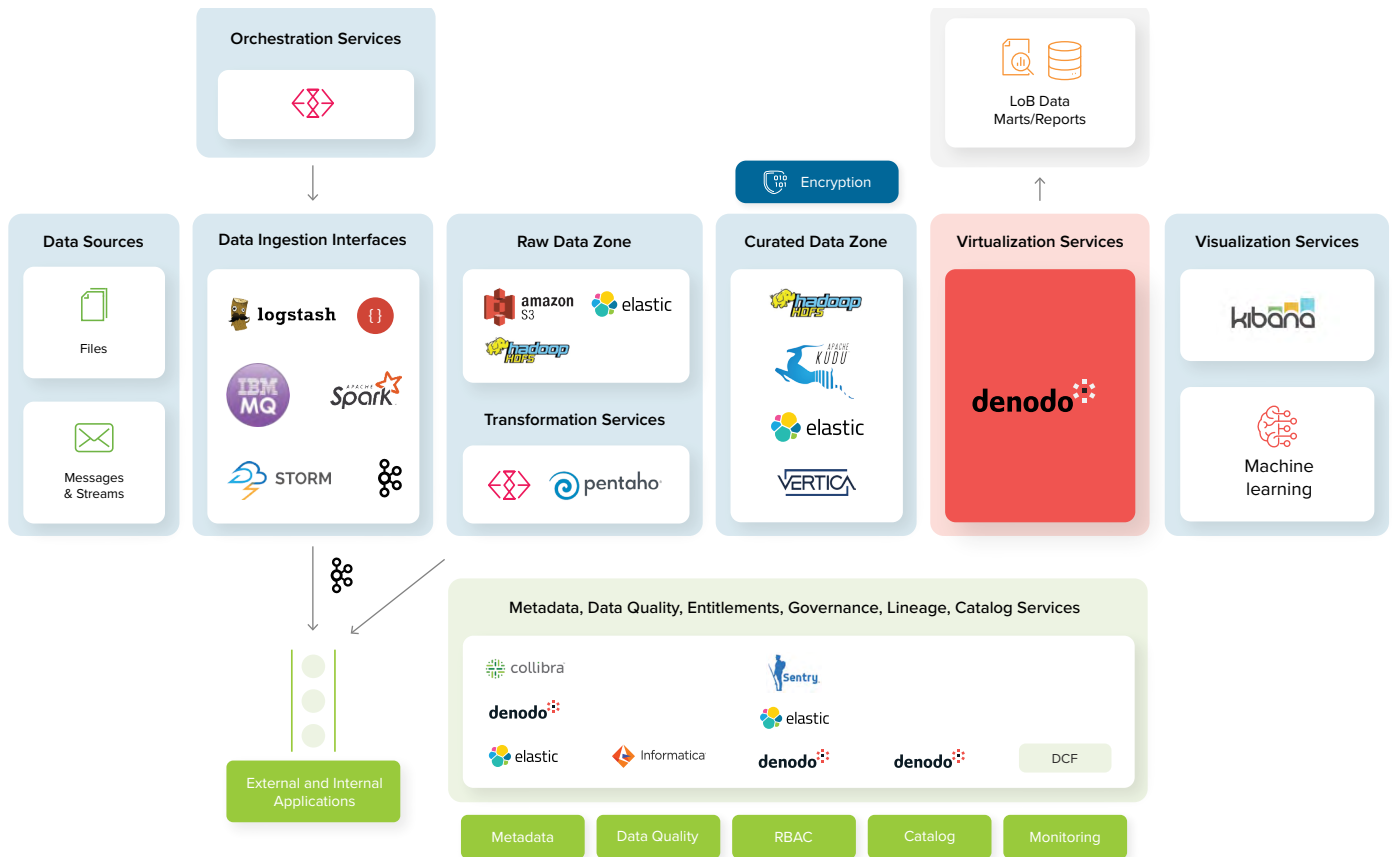
Benefits:

By implementing the Denodo Platform, the bank's EDP was able to meet all of its stringent requirements. In particular:

- The data virtualization layer has made data consumption fairly easy and straightforward. Any application that wishes to consume data simply connects to the data virtualization layer, which goes to any underlying data store and gets the required data.

- As data moves from one application to another across the firm, the Denodo data virtualization layer supports the tracking of data lineage.

- The data virtualization layer supports bank's enforcement of RBAC policies, through the RBAC API, to easily manage authentication.

- The bank found the data virtualization layer to be especially useful in clearing markets, where there is a need to join reference data and FX (foreign exchange) data for ad-hoc data analysis.

- In the asset management space, the data virtualization layer facilitates the joining of SQL Server data with the data in Excel spreadsheets, which is generated by operational personnel.