Logical Data Lake

The expanding volume and variety of data originating from sources that are both internal and external to the enterprise are challenging businesses in harnessing their big data for actionable insights. In their attempts to overcome big data challenges, organizations are exploring data lakes as consolidated repositories of massive volumes of raw, detailed data of various types and formats. But creating a physical data lake presents its own hurdles, one of which is the need to store the data twice which can lead to governance challenges with regard to data access and quality. Also, data lakes can become data silos since they are often built to target particular departments, such as Marketing, and subsequently must be combined with other enterprise data (e.g., CRM, ERP, or other data lakes) for analysis.

Limitations of Physical Data Lakes

Physical data lake solutions are constrained by factors such as duplication of data, governance limitations, and the risk of becoming another data silo.

- Implementation of physical data lakes necessitates storing and maintaining two sets of data, one resident in the source system and the other resident in the data lake. Such duplication of data storage and maintenance incurs greater expense compared to that for a single set of data. Furthermore, maintaining duplicate sets of data creates data quality challenges due to the need to update the data in two distinct locations.

- Governance of the large amounts of data characteristically stored in physical data lakes can be challenging. Restrictions on data access must be maintained as existing data is moved into the data lake and as new data is created within the data lake.

- Data lakes often are built for specific departments, with the potential for creating silos where the data is isolated from other corporate data. The siloed data must then be "discovered" and integrated with the other enterprise data in order to provide a "complete picture" that is necessary to make informed decisions.

SOLUTION
Logical Data Lake

INDUSTRY
Applicable to all Industries
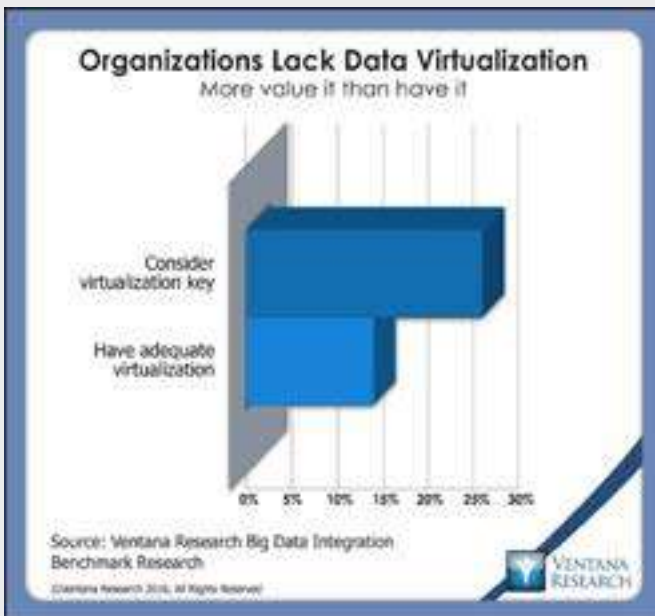
WEBSITE
www.denodo.com

PRODUCT OVERVIEW
Denodo Platform offers the broadest access to structured and unstructured data residing in enterprise, big data, and cloud sources, in both batch and real-time, exceeding the performance needs of data-intensive organizations for both analytical and operational use cases, delivered in a much shorter timeframe than traditional data integration tools.

Organizations Lack Data Virtualization
More value it than have it

Consider virtualization key
Have adequate virtualization

0% 5% 10% 15% 20% 25% 30%

Source: Ventana Research Big Data Integration Benchmark Research

(Ventana Research 2016, All Rights Reserved)

VENTANA RESEARCH

Organizations Lack Advanced Data Virtualization Capabilities.

One-fourth (26%) of organizations participating in our big data integration benchmark research said that data virtualization is a key activity for big data analytics. However, only 14% said that they have completely adequate data virtualization capabilities, revealing a gap between opinion and implementation.

Overcoming the Limitations of Physical Data Lakes

Data virtualization technology provides an agile and cost-effective approach to combining, governing, and managing data in data lakes, and to overcoming the inherent challenges presented by physical data lake silos.

Implement a Single Logical Data Lake Using Data Virtualization.

Improves the enterprise functionality of data lakes by combining one or more physical data lakes with other enterprise data.

Provides a way to access data from separate systems through an abstraction layer that makes it appear as if the data were in a single data lake.

Improves an organization's ability to govern and extract more value from its data lakes by extending them as logical data lakes.
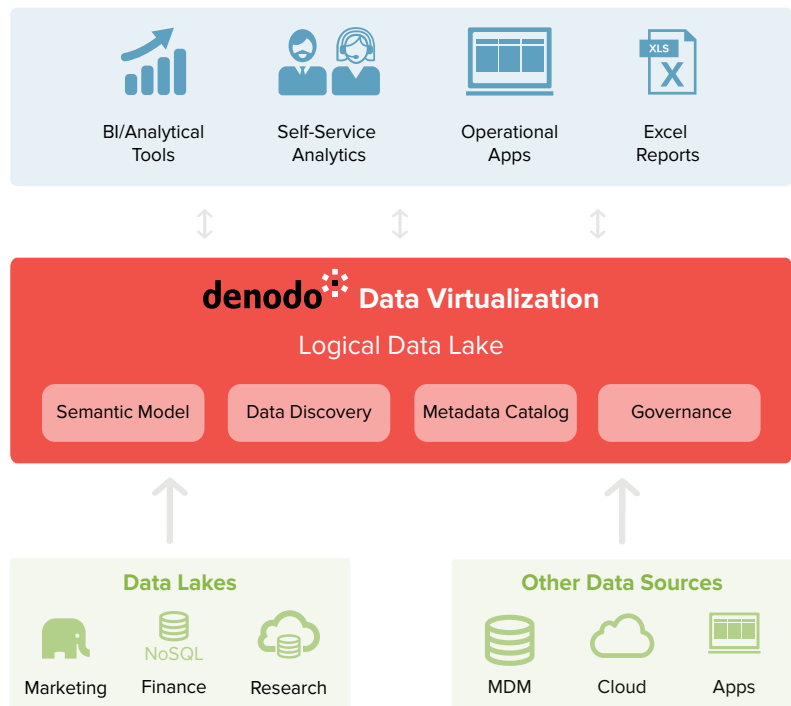
Solution Highlights

To overcome the limitations of physical data lakes, progressive organizations are turning to data virtualization to extend their physical data lakes by creating a "virtual" or "logical" data lake through a layer of abstraction.

Data virtualization can facilitate and expedite accessing and exploring critical data in a cost-effective manner, and assist organizations in deriving more value from their data lakes and their information.

Data virtualization technologies can improve an organization's ability to govern and extract more value from its data lakes by extending them as logical data lakes.

A Single Governed Logical Data Lake

Data Virtualization combines one or more physical data lakes with other enterprise data to create a "virtual" or "logical" data lake.

BI/Analytical Tools · Self-Service Analytics · Operational Apps · Excel Reports

denodo **Data Virtualization**

Logical Data Lake

Semantic Model · Data Discovery · Metadata Catalog · Governance

Data Lakes
Marketing · NoSQL Finance · Research

Other Data Sources
MDM · Cloud · Apps

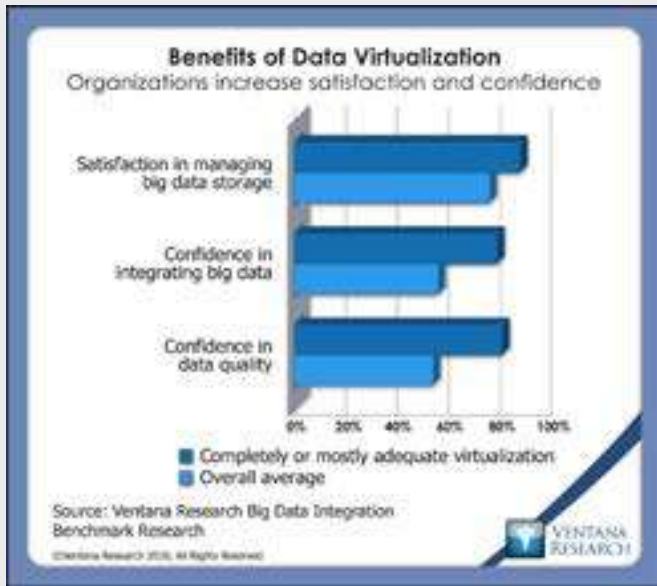## Denodo Platform Bridges Distinct Data Architectures.

Data virtualization bridges different architectures such as data lake, traditional data warehouse and others. This enables organizations to retain existing solution investments while at the same time modernize the data architecture to support new requirements in an agile manner.

Benefits of Managing Data Lakes

Benefits of Data Virtualization
Organizations increase satisfaction and confidence

Source: Ventana Research Big Data Integration Benchmark Research

Data Virtualization is a Practical Strategy for Managing Data Lakes.

Data virtualization imparts agility to big data endeavors. It enables users to quickly combine sources of information without needlessly spending time installing and configuring new databases or clusters to store the consolidated information. The reduction in effort and cost to combine and access information promotes greater exploration of the data which can lead to new insights. Data virtualization eliminates the cost of storing information twice and the need to update information in multiple places since information is not duplicated. Data virtualization prevents the formation of data silos from physical data lakes. Data virtualization is the most viable solution for managing physical data lakes, and providing access to all the information contained therein for enabling an organization to power its analytics.

Case Studies

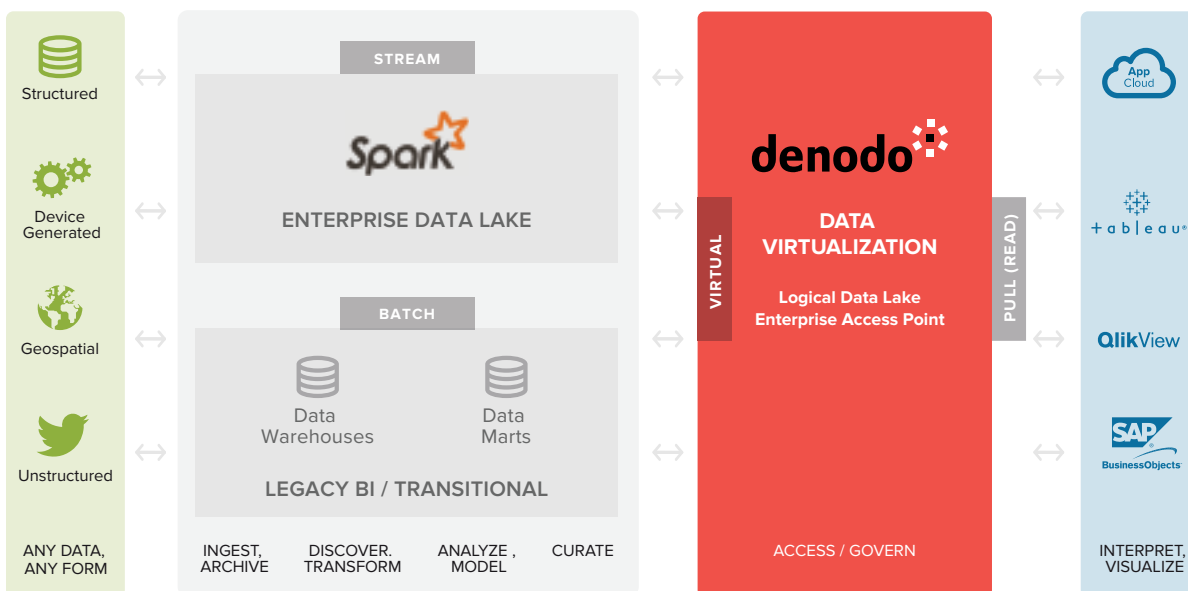## Logical Data Lake Case Study: Autodesk.

Autodesk is a multinational software corporation that develops software for the architecture, engineering, construction, manufacturing, media and entertainment industries. Autodesk uses a logical data lake approach, powered by data virtualization, to complement both its existing business intelligence platform and its operational enterprise data warehouse reporting model.

Benefits:

Autodesk successfully changed their business revenue model from a conventional perpetual license model to subscription-based license model.

For the first time, Autodesk enabled single-point security enforcement and a uniform data layer for access.

Autodesk uses data virtualization layer to combine data from big data and Cloud sources with other enterprise data to provide unified data access to consuming applications in an agile manner.

Logical Data Lake Case Study: a Large Machinery and Equipment Firm.

A leading Construction Manufacturer Improves Service Delivery and Revenue using data virtualization. A logical data lake approach using data virtualization combines sensor data from equipment with parts, maintenance, and dealer information from traditional systems in support of analytical and operational needs.
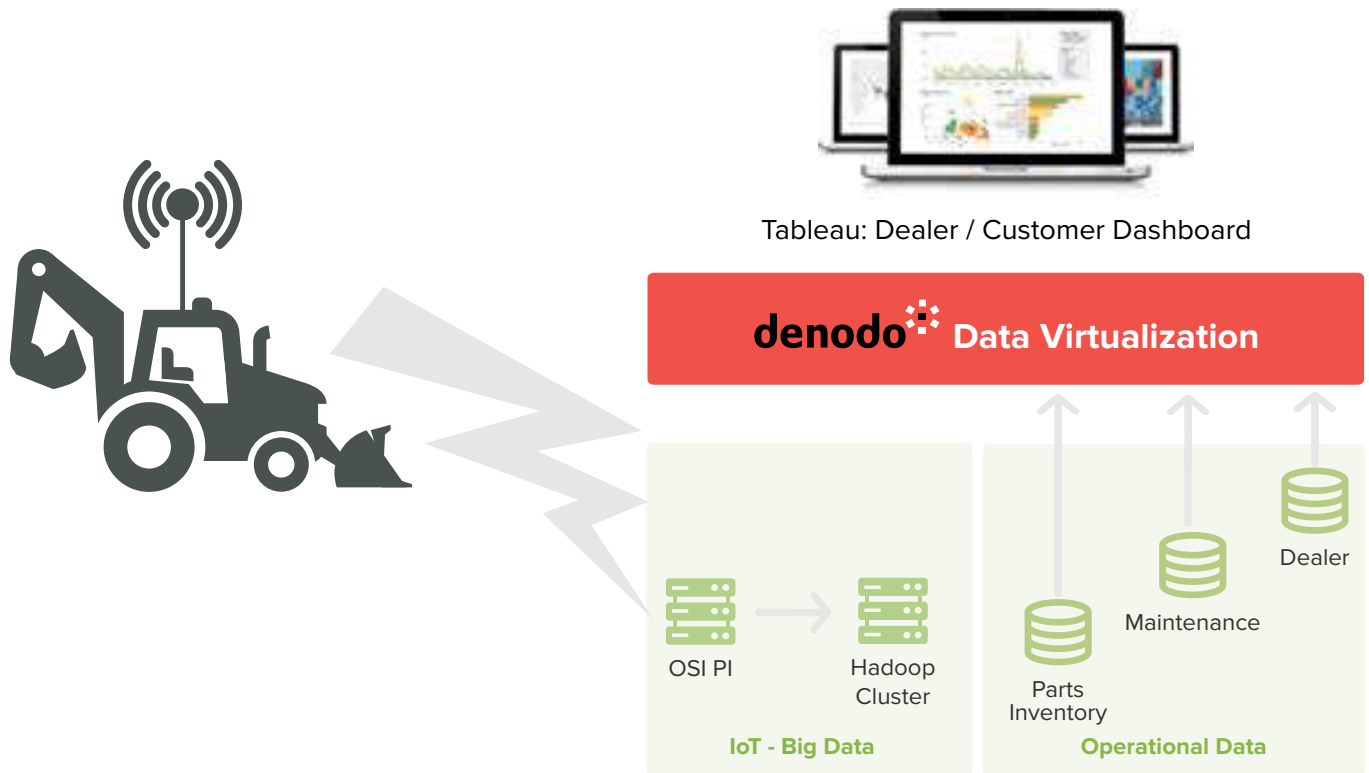
Benefits:

Improved asset performance and proactive maintenance.

Reduced warranty costs due to proactive maintenance of parts preventing parts failure.

Optimized pricing for services and parts among global service providers.

Increased revenue from sale of services and parts. New Business Model opportunities based on real-time analysis of detailed sensor data.

Tableau: Dealer / Customer Dashboard

denodo Data Virtualization

OSI PI → Hadoop Cluster

IoT - Big Data

Parts Inventory

Maintenance

Dealer

Operational Data

---

denodo

About Denodo
Denodo is the leader in data virtualization providing agile, high performance data integration and data abstraction across the broadest range of enterprise, cloud, big data and unstructured data sources, and real-time data services at half the cost of traditional approaches. Denodo's customers across every major industry have gained significant business agility and ROI.

For more information, visit www.denodo.com or call +1 877 556 2531 / +44 (0) 20 7869 8053.