



Enterprise Data Management
DEMO DAY

Denodo Data Virtualization and Data Integration Update

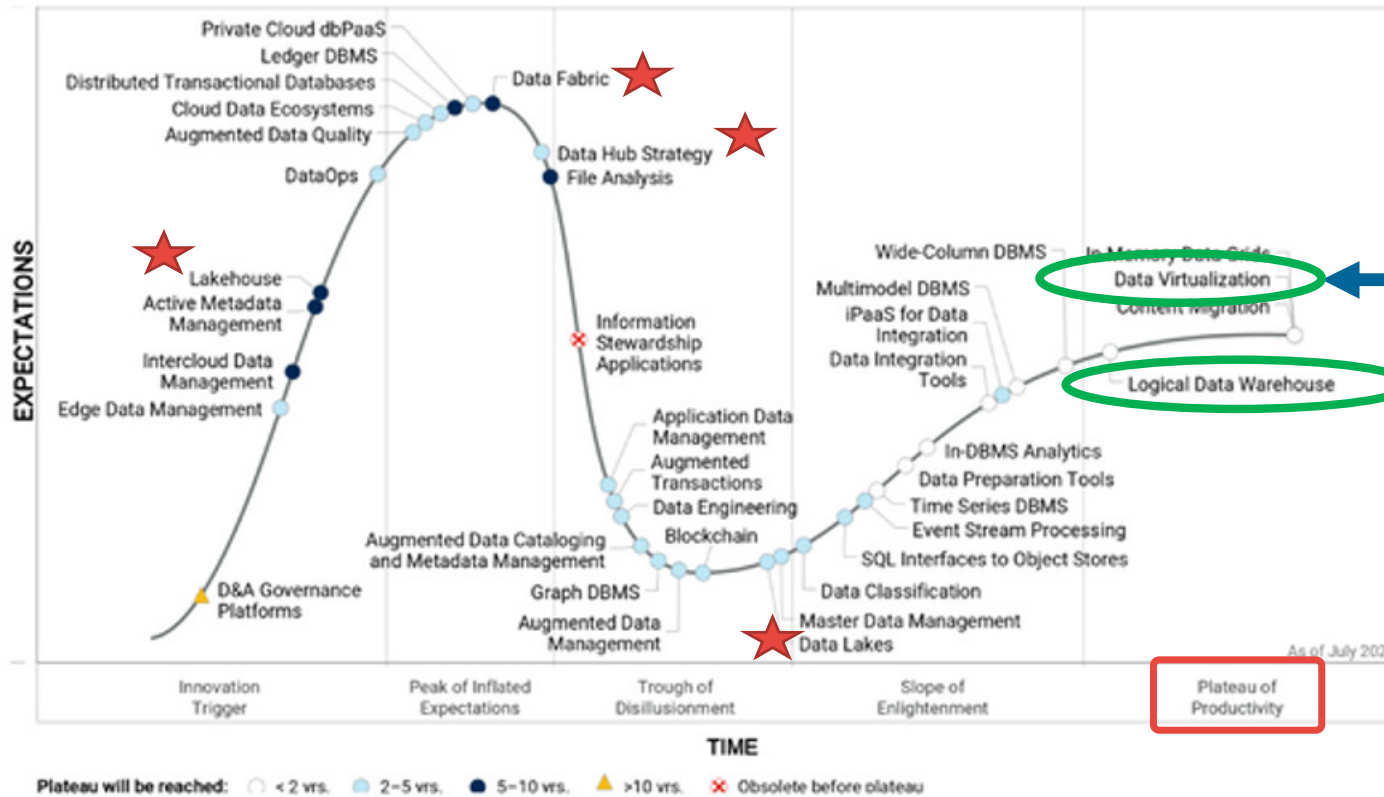
*Building a Modern Enterprise Data Architecture with
AI and Active Metadata*

Dave Chiou - Senior Solutions Engineer
dchiou@denodo.com



2021 Gartner Data Management Hype Cycle

Hype Cycle for Data Management, 2021



“Data Virtualization can be deployed with **low risk and effort** to achieve **maximum value**.”

Data Virtualization is a key capability for **Logical Data Warehouse, Data Fabric, Data Hub, Microservices Architectures**

Source: Gartner (July 2021)

747543

Source: Gartner (July 2021)

2021 Gartner MQ Data Integration – Denodo Leader



“ Denodo enables its customers to **connect distributed data** through **business-friendly semantic models** that **decouple data from its location and physical schemas**. It reflects a longtime focus and strength in data virtualization that **enables agile data integration and delivery.**”

-- Gartner Magic Quadrant for Data Integration Tools, August 2021

- Denodo “ 2nd Fastest Growing DI Vendor” among Leaders
- **Active Metadata** is a “significant driver of market vision”
- Data Fabric is critical and **Automation** using **ML/AI** is key




Gartner

Source: Gartner (August 2021)



Data Virtualization, ETL, ESB Compared

Guiding principles on when to use DV versus other methods for data delivery are driven by your non-functional requirements (use cases, TCO, time-to-market)

	Physical Movement and Consolidation		Logical Abstraction and Virtual Integration	Synchronization and Propagation
				
MIDDLEWARE	ETL	CDC	Data Virtualization	EAI / ESB
PURPOSE	DB ↔ DB	DB ↔ DB	DB ↔ Applications	Application ↔ Application
MODEL	Scheduled	Event Driven	On demand	Event Driven
STRENGTHS	<ul style="list-style-type: none"> Building DWs and MDM Hubs Complex workflows and DQ Historical data and cubes 		<ul style="list-style-type: none"> Distributed access and delivery Agility and timeliness Logical Data Warehouse 	<ul style="list-style-type: none"> Business process automation Transaction propagation Messaging with small payloads

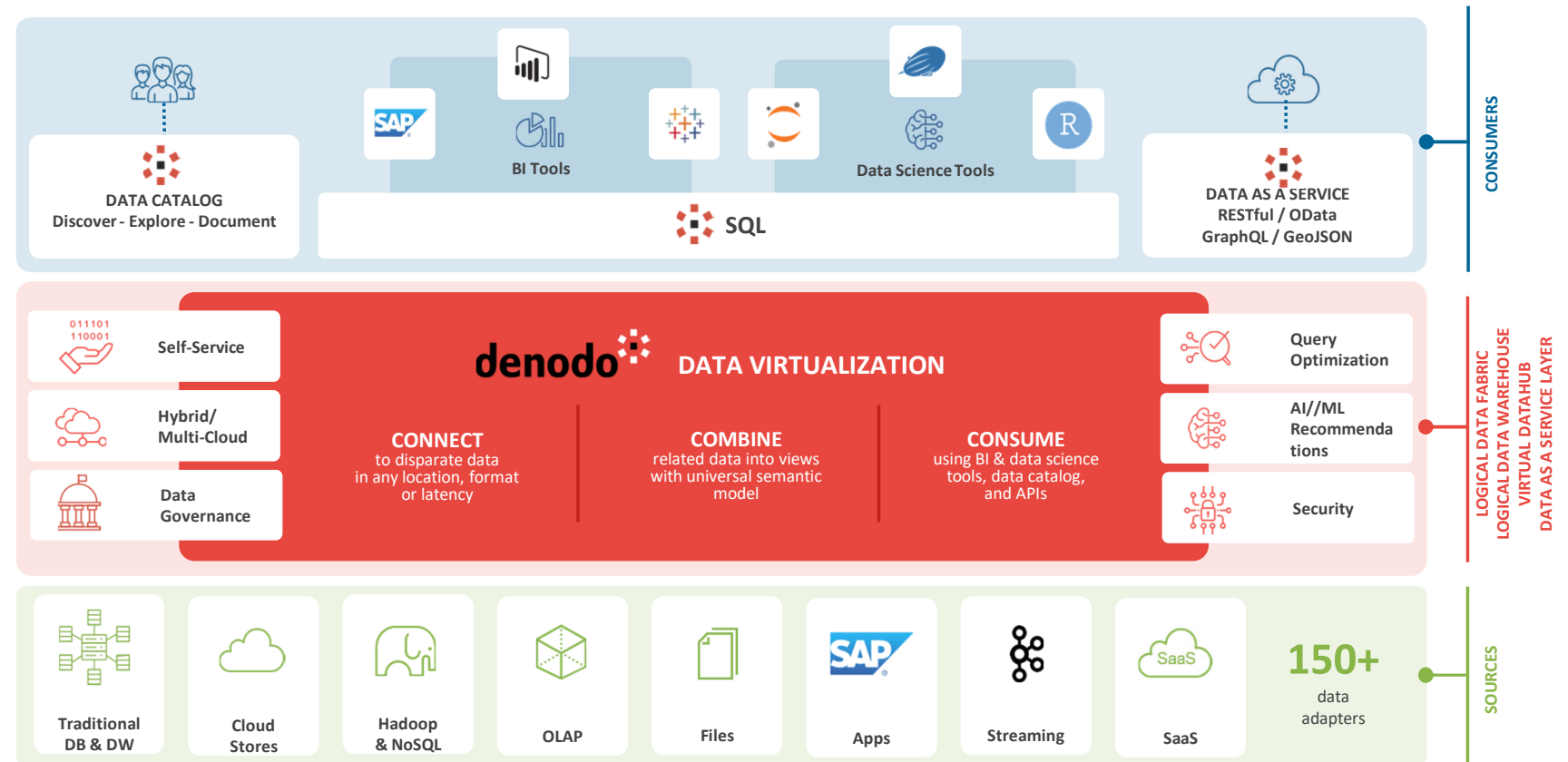
Logical Data Fabric

Logical Data Fabric is the evolution of Denodo's Data Virtualization Logical Architecture

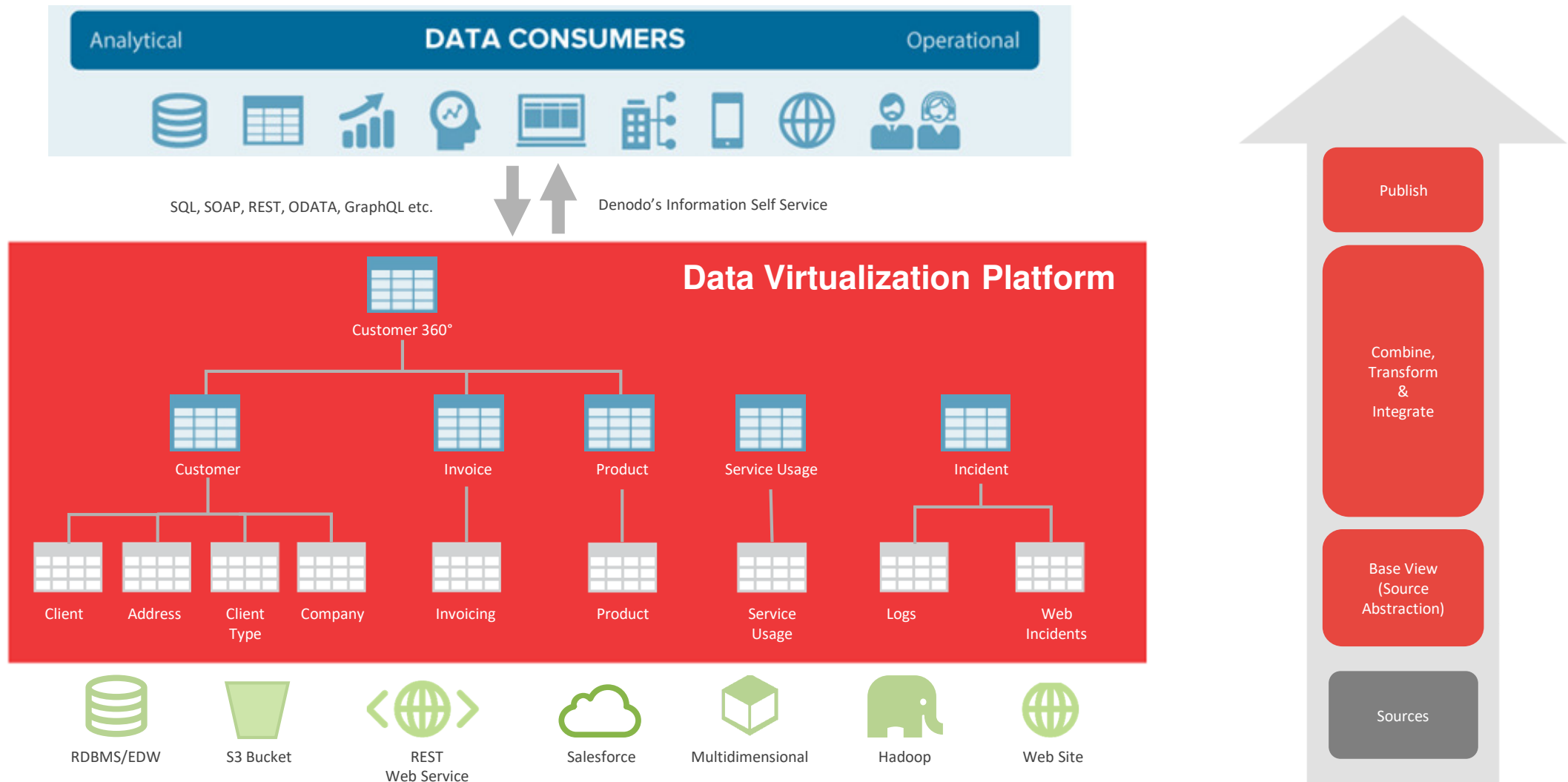
Gartner

Demystifying the Data Fabric,
September 2020

The core of the matter is being able to **consolidate many diverse data sources** in an efficient manner by allowing **trusted data to be delivered from all relevant data sources to all relevant data consumers** through **one common layer**.



How Does It Work?



Automatic Recommendations for Smart Query Acceleration

Denodo uses Artificial Intelligence to automatically recommend selective materialization of datasets (summaries) to increase performance

- AI Algorithms using Active Metadata
 - Usage history from Denodo monitor
 - Data profiling statistics of sources
 - Cost simulations to generate summary recommendations
- Optimization
 - Generic enough to cover multiple queries
 - Specific enough to keep it small and fast

Workload preferences:

- Minimum query duration (ms): 3000
- Maximum number of queries to consider: 500
- Initial date to consider: to_localdate('dd/MM/yyyy','6/4/2021')
- Last date to consider: to_localdate('dd/MM/yyyy','6/4/2021')

Summary candidates preferences:

- Maximum size for summary candidates (number of rows): 200000
- Maximum number of summary candidates: 10
- Recommendations time limit (sec): 1800
- Database for recommended summaries: query_acceleration
- Folder for recommended summaries: /
- Ignore existing summaries
- Preferred views: + None
- Preferred target data sources: + None
- Excluded target data sources: + None

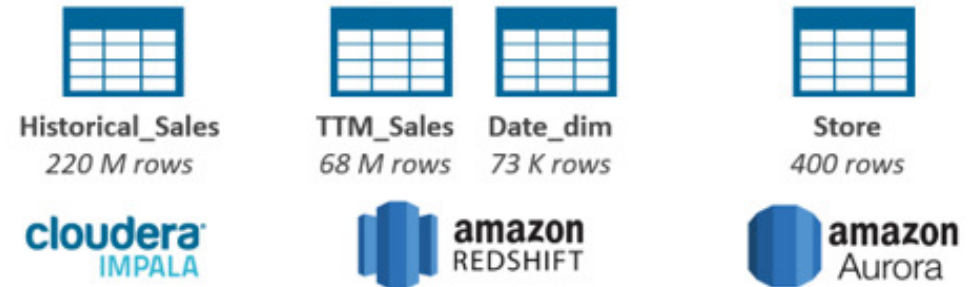
[Generate recommendations](#)

The process has finished successfully.
Execution time: 7 sec
Number of queries satisfying the selected criteria: 15
Number of summary recommendations: 5

	Name	Data source	Views	Estimated size (r...	Usage (Num. que...	Absolute gain fac...	Relative gain fac...
<input type="checkbox"/>	sum1-redshift_tpcd...	redshift_tpcds red...	redshift_tpcds stor...	409	4	1.54	1.54
<input type="checkbox"/>	sum2-redshift_tpcd...	redshift_tpcds red...	redshift_tpcds dat...	8301	4	1.1	1.17
<input checked="" type="checkbox"/>	sum3-redshift_tpcd...	redshift_tpcds red...	redshift_tpcds stor...	73658	6	2.0	1.11
<input type="checkbox"/>	sum4-redshift_tpcd...	redshift_tpcds red...	redshift_tpcds dat...	178	2	1.08	1.0
<input type="checkbox"/>	sum5-redshift_tpcd...	redshift_tpcds red...	redshift_tpcds stor...	37	2	1.35	1.0

Query Acceleration with Aggregate Awareness

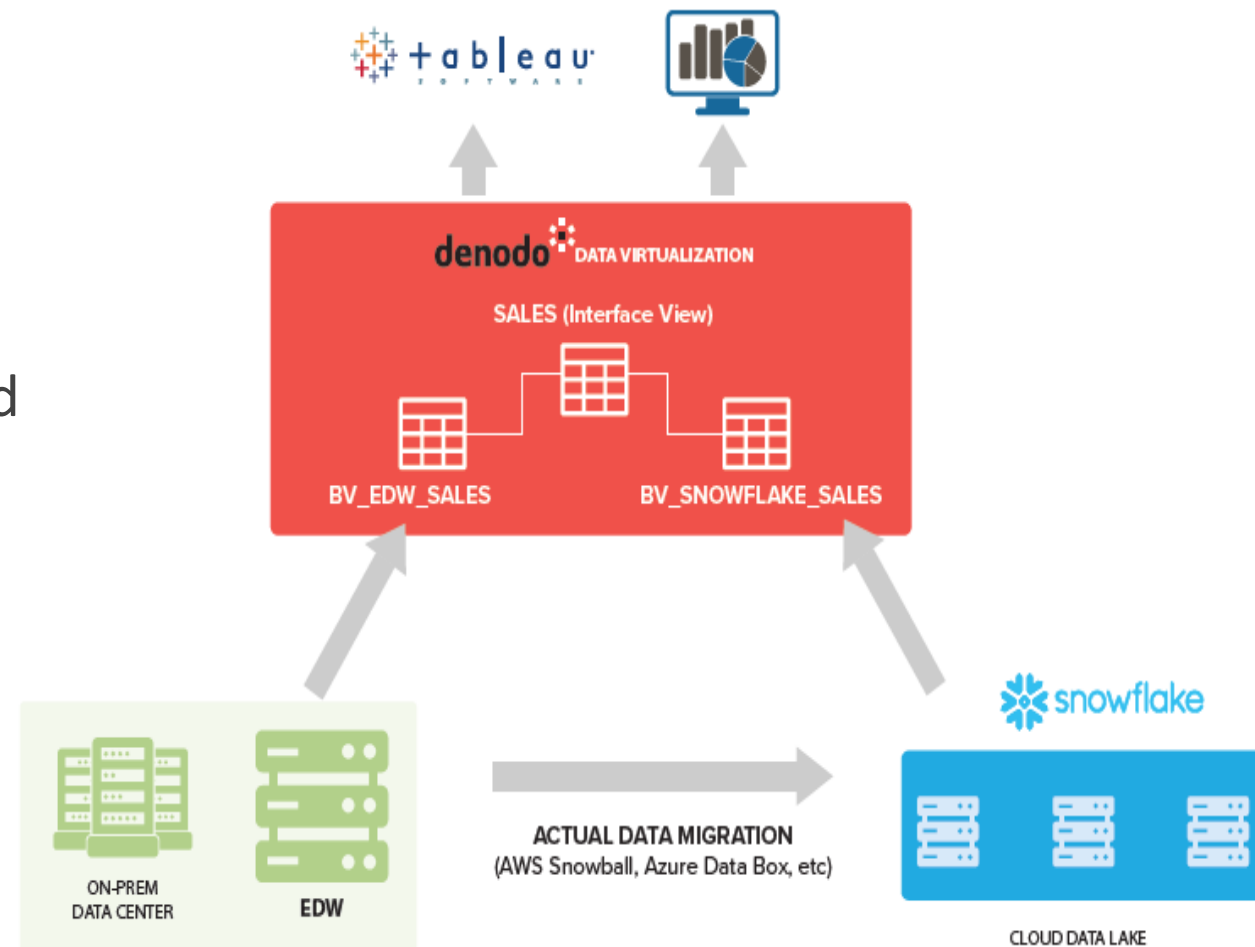
- TPCS-DS data:
 - Distributed in 3 different systems
 - Tables with hundreds of millions of rows
- Summary: total sales by store id, sold_date_id



Query	Execution Time (no acceleration)	Execution Time (acceleration)	Performance Gain	Summary used
Total sales by year	15.45 s	2.38 s	6.5 x	summary_total_by_store_day
Total sales by quarter, store name and city	22.49 s	2.62 s	8.57 x	summary_total_by_store_day
Total sales by store and City for last quarter	14.71 s	0.47 s	31.1 x	summary_total_by_store_day
Total sales in a specific store	14.36 s	2.66 s	5.39 x	summary_total_by_store_day
Total sales in a specific store and year	14.32 s	3.18 s	4.0 x	summary_total_by_store_day

Data/Cloud Migration in the Data Fabric

- Large or critical Cloud migrations are risky
 - Big Bang approach is not advised
- Phased approach is recommended
 - Select data set to migrate, copy to Cloud
 - Test and tune data access, then go live
 - Repeat for next data set and so on
- Use Denodo as abstraction layer during migration process
 - Isolate users from shift of data





Demonstrations

December 7, 2020

Building the Data Virtualization Layer
User Access to Data Virtualization Layer
Self-Service Data Catalog

September 15, 2021

Logical Data Fabric
AI Recommendation Engine
Data Migration in the Fabric

Key Take-Away



“ Through 2022, **manual data management tasks** will be **reduced by 45%** through the addition of **machine learning and automated service-level management.**”

-- Gartner Magic Quadrant for Data Integration Tools, August 2021

- Benefits
 - Data Virtualization is mature today and positions you for emerging architectures
 - Manage performance and query recommendations through AI and Active Metadata
 - Quickly delivers data to users and business units
 - Lowers costs, increases productivity and systems learn automatically
 - Agility, Flexibility, Speed, Rapid prototyping
 - Abstraction layer – Migration to the cloud, Decouple and Futureproof your IT architecture
 - Enable user self-service to all your data in a secure and governed manner

Denodo Resources

www.denodo.com

- Denodo DataFest 2021 Conferences

- AMERICAS: Oct 6-7
- APAC: Oct 13-14
- EMEA: Oct 20-21



- Denodo Test Drives on AWS, Azure and GCP

- Denodo Express download

- Data Virtualization for Dummies

- Denodo Demo: Dec 7, 2020

- <https://www.dataversity.net/denodo-demo-data-virtualization-and-data-integration/>

Thanks!

Q&A