IT Modernization
Logical Data Architecture
Using ML / AI

DIA CIO Virtual Summit 4/27/21

Presenter: Richard Walker,

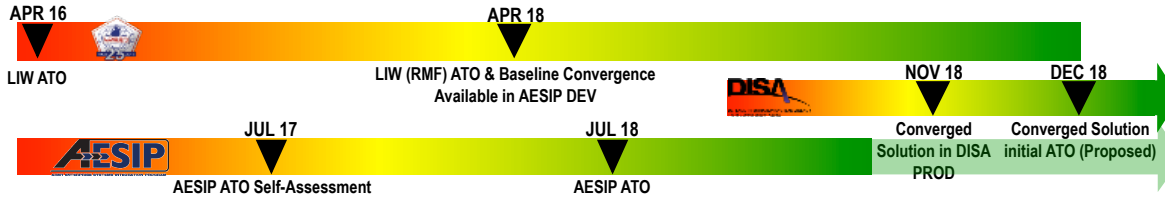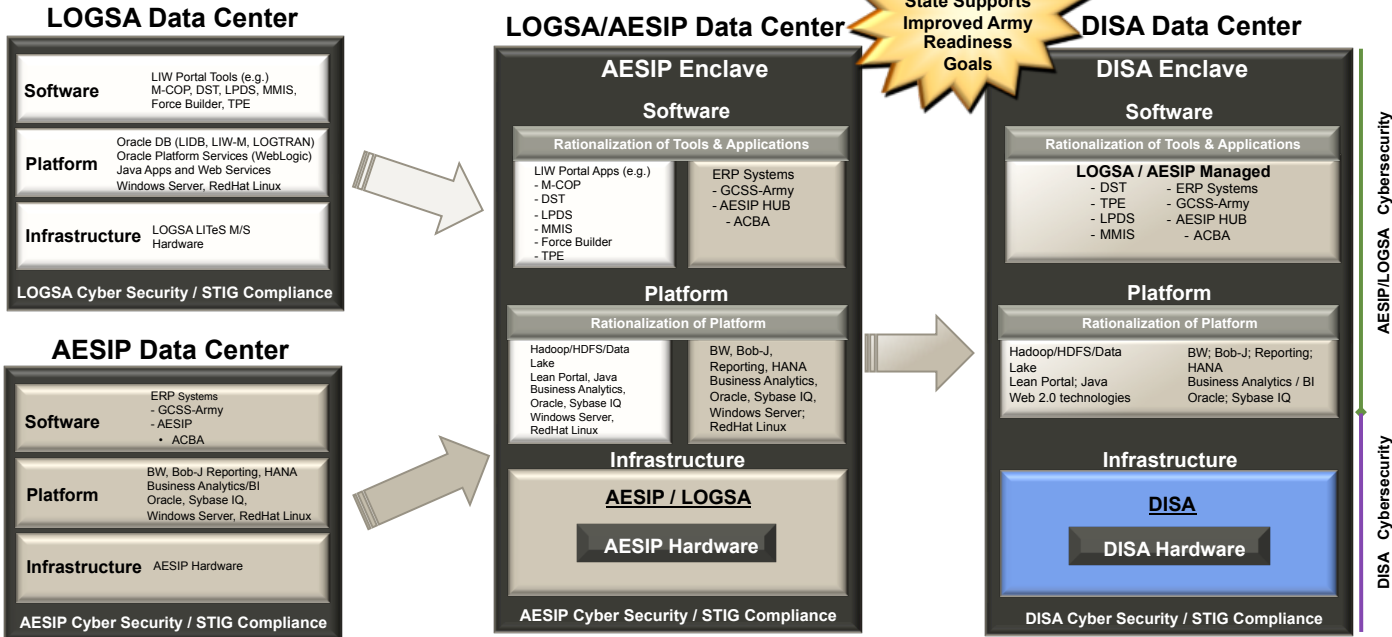Chief Customer Officer

rwalker@denodo.com

denodo

Logical Data Layers - now a Standard Gov. Arch.

- **DOD - DIA**
  - **MARS/TALOS - Oper. & BI Analytics**
  - **Data Hub**
  - **SITE III**
  - **Digital Transf. of Found. Mil. Intell. (MTMI)**
  - **Military Intelligent Data Base (MIDB)**

- **DOD**
  - **Off of Comptroller** – Self-Service Agile BI Analytics
  - **Joint Strike Fighter (JSF)** – Digital Transformation, Multi-cloud (AWS & Azure)
  - **Joint Artificial Intelligence Center** - governance semantic layer

- **US Space Force** - digital engineering platform & Secure Data Fabric

- **US Cyber Command** –use AI/ML to identify threats in networks

- **US CENTCOM** – Records Mng / Cloud

- **US SOCOM** – SOF Data Layer

- **US Army**
  - **Material Command** – Data Services Layer
  - **Corp of Engineers** – Enter. Virtual Viewer (EVV)
  - **Futures Command / C5ISR** – Unified Data Ecosystem

- **USAF**
  - **CDO Data VAULTIS** – Data access, governance and semantic layer
  - **Data Access Layer for Data Lakes and Multi-Cloud**
  - **Adv Battle Mng System** –
  - **Research Laboratory (AFRL)** Data Scientist Ecosystem

- **Missile Defense Agency** – Logical data layer

- **US Navy**
  - **Navy Data Platform**
  - **Navy Sea Systems Command** – Universal Hub for Big Data

- **DOE NNSA (Nat'l Nuclear Security Admin)** – ent. wide governance & security

denodo

US Army Material Command - Convergence/Modernization

LOGSA | AESIP | DISA

Operational End State Supports Improved Army Readiness Goals

LOGSA Data Center

Software	LIW Portal Tools (e.g.) M-COP, DST, LPDS, MMIS, Force Builder, TPE
Platform	Oracle DB (LIDB, LIW-M, LOGTRAN) Oracle Platform Services (WebLogic) Java Apps and Web Services Windows Server, RedHat Linux
Infrastructure	LOGSA LITeS M/S Hardware

LOGSA Cyber Security / STIG Compliance

AESIP Data Center

Software	ERP Systems - GCSS-Army - AESIP • ACBA
Platform	BW, Bob-J Reporting, HANA Business Analytics/BI Oracle, Sybase IQ, Windows Server, RedHat Linux
Infrastructure	AESIP Hardware

AESIP Cyber Security / STIG Compliance

LOGSA/AESIP Data Center

AESIP Enclave

Software

Rationalization of Tools & Applications

LIW Portal Apps (e.g.) - M-COP - DST - LPDS - MMIS - Force Builder - TPE	ERP Systems - GCSS-Army - AESIP HUB - ACBA

Platform

Rationalization of Platform

Hadoop/HDFS/Data Lake Lean Portal, Java Business Analytics, Oracle, Sybase IQ Windows Server, RedHat Linux	BW, Bob-J, Reporting, HANA Business Analytics, Oracle, Sybase IQ, Windows Server; RedHat Linux

Infrastructure

AESIP / LOGSA

AESIP Hardware

AESIP Cyber Security / STIG Compliance

DISA Data Center

DISA Enclave

Software

Rationalization of Tools & Applications

LOGSA / AESIP Managed

- DST - TPE - LPDS - MMIS	- ERP Systems - GCSS-Army - AESIP HUB - ACBA

Platform

Rationalization of Platform

Hadoop/HDFS/Data Lake Lean Portal; Java Web 2.0 technologies	BW; Bob-J; Reporting; HANA Business Analytics / BI Oracle; Sybase IQ

Infrastructure

DISA

DISA Hardware

DISA Cyber Security / STIG Compliance

AESIP/LOGSA Cybersecurity

DISA Cybersecurity

Timeline:

- APR 16 — LIW ATO
- APR 18 — LIW (RMF) ATO & Baseline Convergence Available in AESIP DEV
- NOV 18 — Converged Solution in DISA PROD
- DEC 18 — Converged Solution initial ATO (Proposed)

- JUL 17 — AESIP ATO Self-Assessment
- JUL 18 — AESIP ATO

Logical Data Services Layer

- **The Organization wants:**

- One Place for Users to go and get any Data

- Results Presented in a Useable Format

- Can Access & View results from any tool / data source

Logical DS Layer should provide:

- **Abstraction** – free users from the underlying complexities of data sources and data types
- **Decoupling** – Decoupling data from the business logic
- **Logical Representation** – User or Enterprise Defined Data Models
- **Real-Time or Right-Time** – Access to data on demand, cached or scheduled batch
- **Lighter, Agile Integration** – Federated views of disparate data; easy to transform, combine, change at will
- **Reduce Replication** – HW, SW and Governance costs of unnecessary data replication
- **Integrates with Existing Tools** – Leverages them in hybrid modes (on-premise and cloud)

U.S. Army Materiel Command LDAC/AESIP Case Study

BI / Reporting Capabilities

- WebLIDB/LMR
- Parts Tracker
- DODAAC Readiness Analyzer
- Customer Satisfaction
- Overage Repairable Management
- Repairable and Recoverable
- Supply Performance Statistics
- Aging Deadline Notifications
- Cost Driver
- DPST
- MMIS
- ARMT
- TPE Planner

Analytics / Visualizations

- Fleet Readiness
- Fleet Projection
- Readiness Improvement Options
- Combat Slant
- CWT and ZPARK
- SSA ASL Performance
- Class IX Parts Impacting Readiness
- Class VII and IX Asset Visibility
- FMC Percentages % by UIC or LIN
- Condition-Based Maintenance (CBM)

Modeling, Simulation and Artificial Intelligence

- LMI DST; COA Analysis, OIB Optimization
- CBM
- ExASL: Stockage determination
- PSCC: Optimization of storage and distribution
- COMPASS: enables Level of Repair Analysis (LORA) studies
- CASA: Life Cycle Cost (LCC)/Total Ownership Cost (TOC) decision support
- Army Air Clearance Authority

Presentation Layer
Semantic / Transformation Layer
Data Foundation Layer
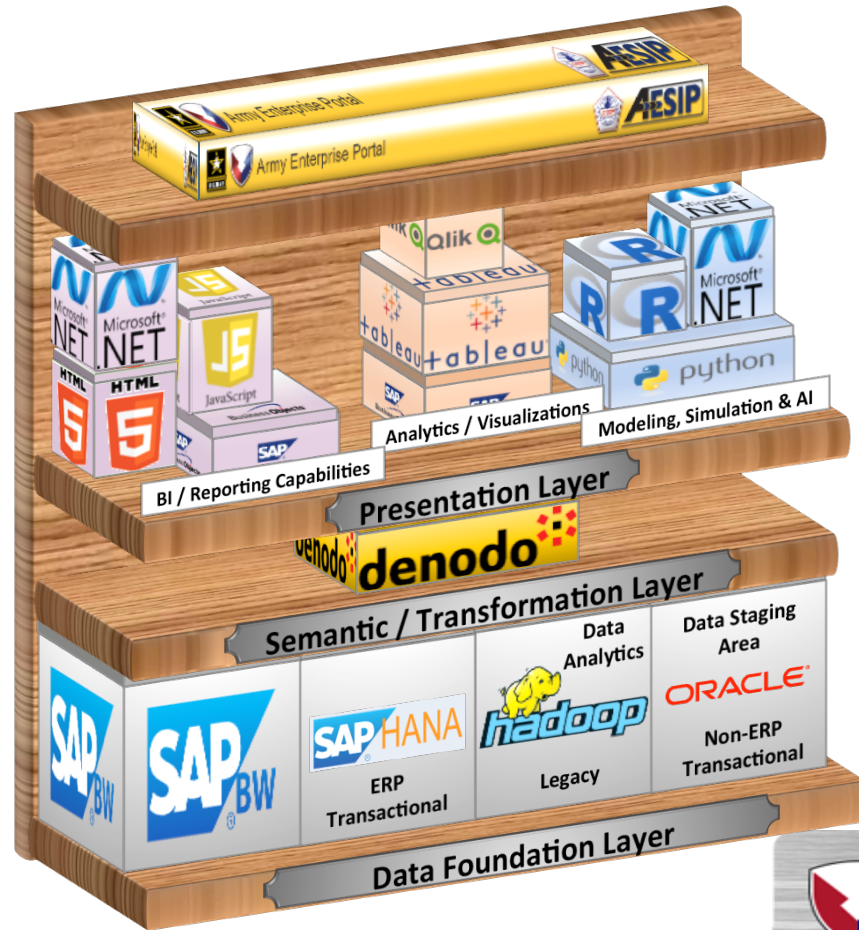
ROI – Creating Services Layer vs Traditional Methods

- **Traditional method – using Oracle tools create extensive materialized database views & publish as web services**

- **Data Virtualization method – create logical database views in DV leaving data in original sources and publish web services.**

- **Performance tests – queries across both methods were within a few % of each other**

- **ROI**

- **97% Faster using Data Virtualization Time**
  - With same staff - completed the same views/ web services in Denodo as in Oracle data stores in less than 3% of the time.
  - Minimal Expertise Req'd with DV - Note that the team used DV without any product training from Denodo whereas Oracle required extensive expertise to build out the views and web services..

- **Saving over $4 MM per year**
  - Due to reduced labor costs, SW/HW storage costs and dramatically shortened timeline for the Modernization/ Convergence project timeline. In fact the data integrations are now completed faster than report writers can keep up with which has never happened before. Traditionally it took weeks or months to complete integration work and now is done in minutes.

- **Security**
  - Follow US Army and NS2 Security checklist for Go-Live Readiness and ATO certifications.
  - US Army LDAC is deployed in the GovCloud.

Apr-23-21

DOD – Missile Defense Agency (MDA)

Overview - MDA HQ Ft. Belvoir, VA. 10k employees, $1.3B budget

- Partof DOD responsible for developing a layered defense against ballistic missiles. 1983 started as Strategic Defense Initiative (SDI), 1993 renamed the Ballistic Missile Defense Organization (BMDO) and 2002 renamed the Missile Defense Agency.

- Ballistic missile defense (BMD) systems are: Patriot PAC-3, Aegis BMD, THAAD, Ground-Based Midcourse Defense system, Multi-Object Kill Vehicle, Kinetic Energy Interceptor and Airborne Laser.   Also funds fundamental research in high-energy physics, supercomputing/computation, advanced materials, and many other science and engineering disciplines.

Use case:

- MDA is a system of systems of war fighting machinery that have to work together – it is incumbent that the operations part of the orchestration that the data flows from them into analytical systems so can see how they work together. Goal is to enable Operating systems working together.

Solution Architecture: Logical Data Layer will:

- **Create a common semantic and governance layer for research analytics of simulations and tests.**

- **Build a virtualized data lake with push down query capability and data cataloging for practicality.**

Benefit

- **Order of magnitude Faster and Easier Analysis of As-Designed vs As-Built vs As-Operate**

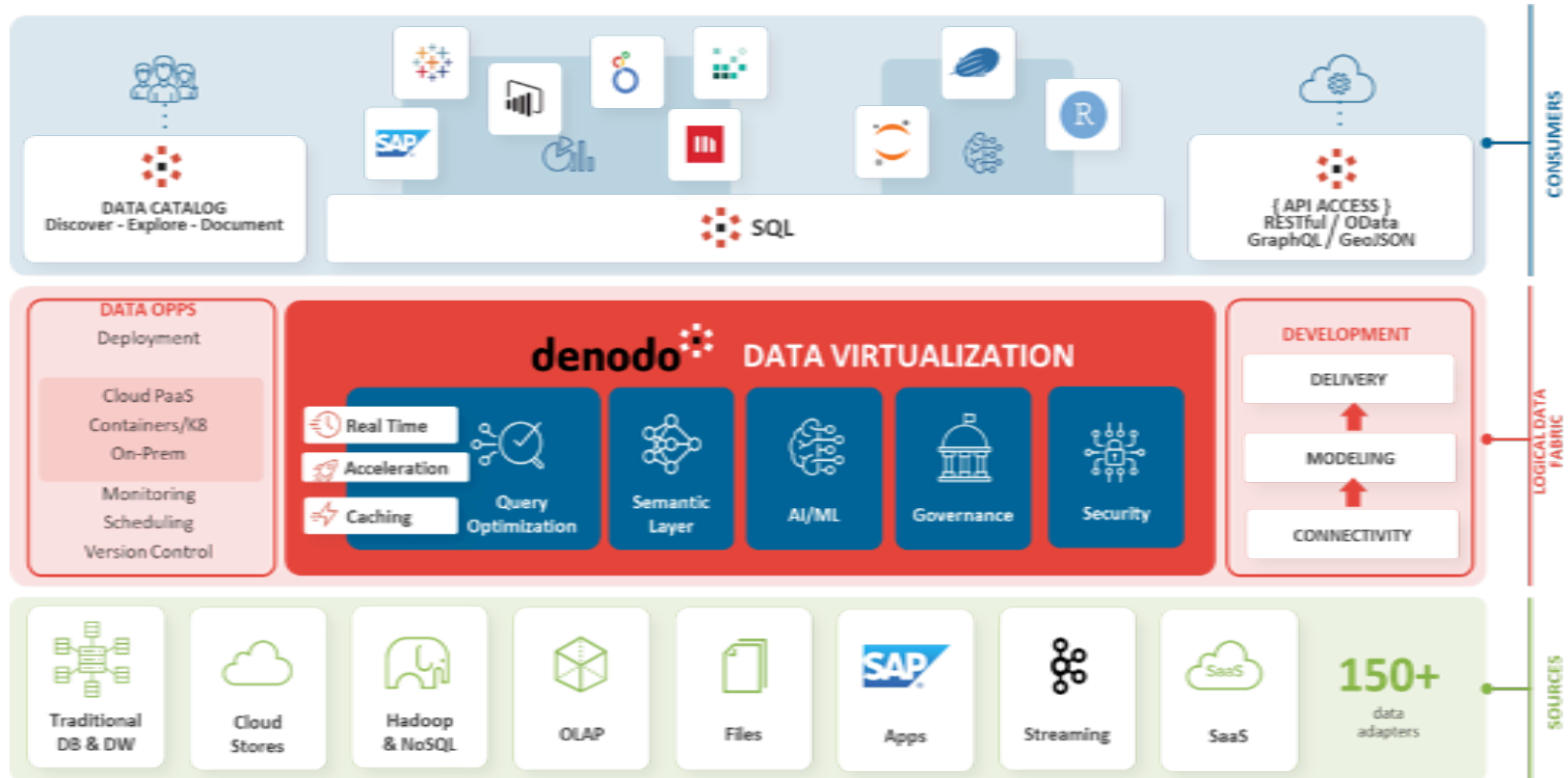- **Then Re-calibrate results back into the Design.**

denodo

DV used as the Governance and Data Security Layer to Share data Across Sites

E.g. Sharing of Eng. & Prod. BOM across sites is now done in real time

DOE's
National Nuclear Security Administration (NNSA) Facility Locations

Lawrence Livermore Nat'l Lab

Nevada National Security Site

Los Alamos Nat'l Lab

Sandia Nat'l Lab

Fermi labs

Kansas City Plant

Y-12 Plant National Security Complex

Naval Nuclear Laboratory

Pantex

Savannah River Site

SECURING AMERICA

Denodo Deployed

4

Logical Data Layer Functional Architecture

DV 14x Faster than Reporting Tools

Reporting Tools are not optimized for federation across sources

```
SELECT c.id, SUM(s.amount) as total
FROM customer c JOIN sales s
ON c.id = s.customer_id
GROUP BY c.id
```

System	Execution Time	Data Transferred	Optimization Technique (automatically selected)
Denodo	9 sec.	1.99 M	Aggregation push-down
Tableau	125 sec.	292.00 M	None: full scan

State OK
Completed yes
SQL sentence
SELECT s0.ren_sum, s0.sum_1, s0.sum_2, s0.product_key FROM (SELECT sum(s1.quantity) AS ren_sum, sum(s1.sale_total) AS sum_1, sum(s1.gross_profit) AS sum_2, s1.product_key AS product_key FROM (SELECT t0.product_key AS product_key, t0.quantity AS quantity, t0.sale_total AS sale_total, t0.gross_profit AS gross_profit FROM sls_historic_sales_fact t0) s1 GROUP BY s1.product_key) s0 WHERE s0.product_key IS NOT NULL
Parameters []
DB URI jdbc:impala://192.168.1.109:21050/globalsales
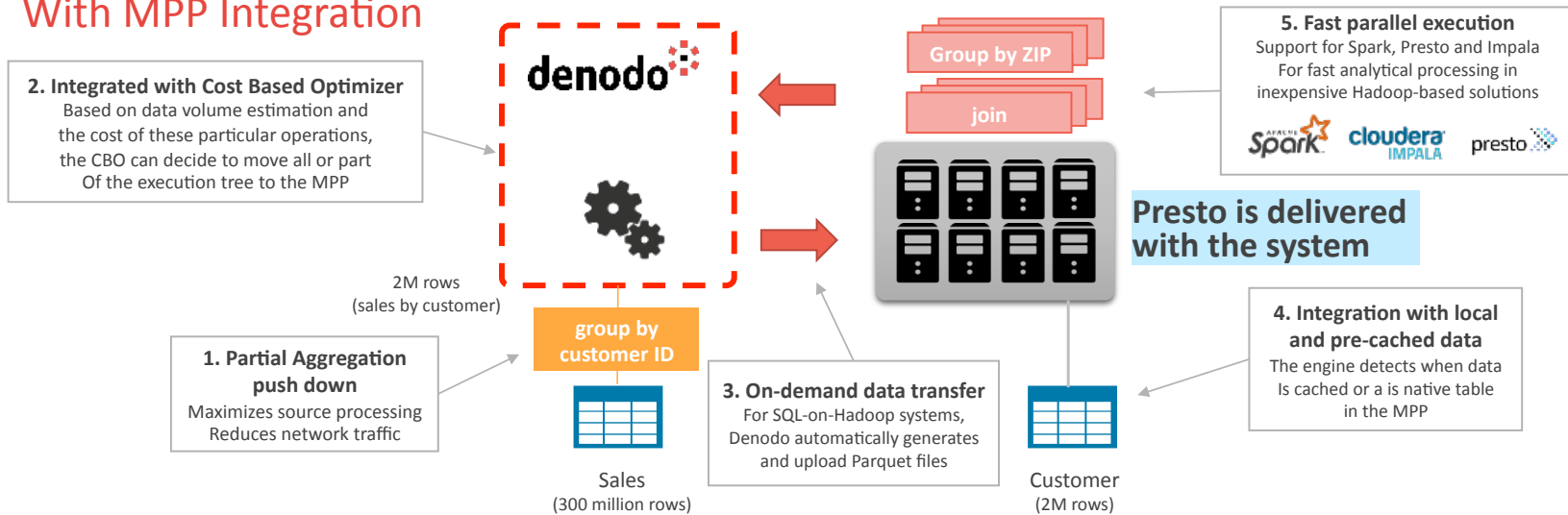User name cloudera

Denodo: aggregation pushdown

Query

SELECT `sls_historic_sales_fact`.`gross_profit` AS `gross_profit`, `sls_historic_sales_fact`.`product_key` AS `product_key`, `sls_historic_sales_fact`.`quantity` AS `quantity`, `sls_historic_sales_fact`.`sale_total` AS `sale_total` FROM `globalsales`.`sls_historic_sales_fact` `sls_historic_sales_fact`

Tableau: full scan

denodo

Massive Parallel Processing:

With MPP Integration

2. Integrated with Cost Based Optimizer
Based on data volume estimation and the cost of these particular operations, the CBO can decide to move all or part Of the execution tree to the MPP

denodo

5. Fast parallel execution
Support for Spark, Presto and Impala
For fast analytical processing in inexpensive Hadoop-based solutions

Spark cloudera IMPALA presto

Group by ZIP

join

Presto is delivered with the system

2M rows
(sales by customer)

1. Partial Aggregation push down
Maximizes source processing
Reduces network traffic

group by customer ID

3. On-demand data transfer
For SQL-on-Hadoop systems, Denodo automatically generates and upload Parquet files

4. Integration with local and pre-cached data
The engine detects when data Is cached or a is native table in the MPP

Sales
(300 million rows)

Customer
(2M rows)

No DV	DV with MPP	% Increase
10 min	11 sec	5455%
14 min	34 sec	2456%
19 min	7.5 min	253%

denodo

Logical Arch. for Multi-Cloud & On-Premise Deployments
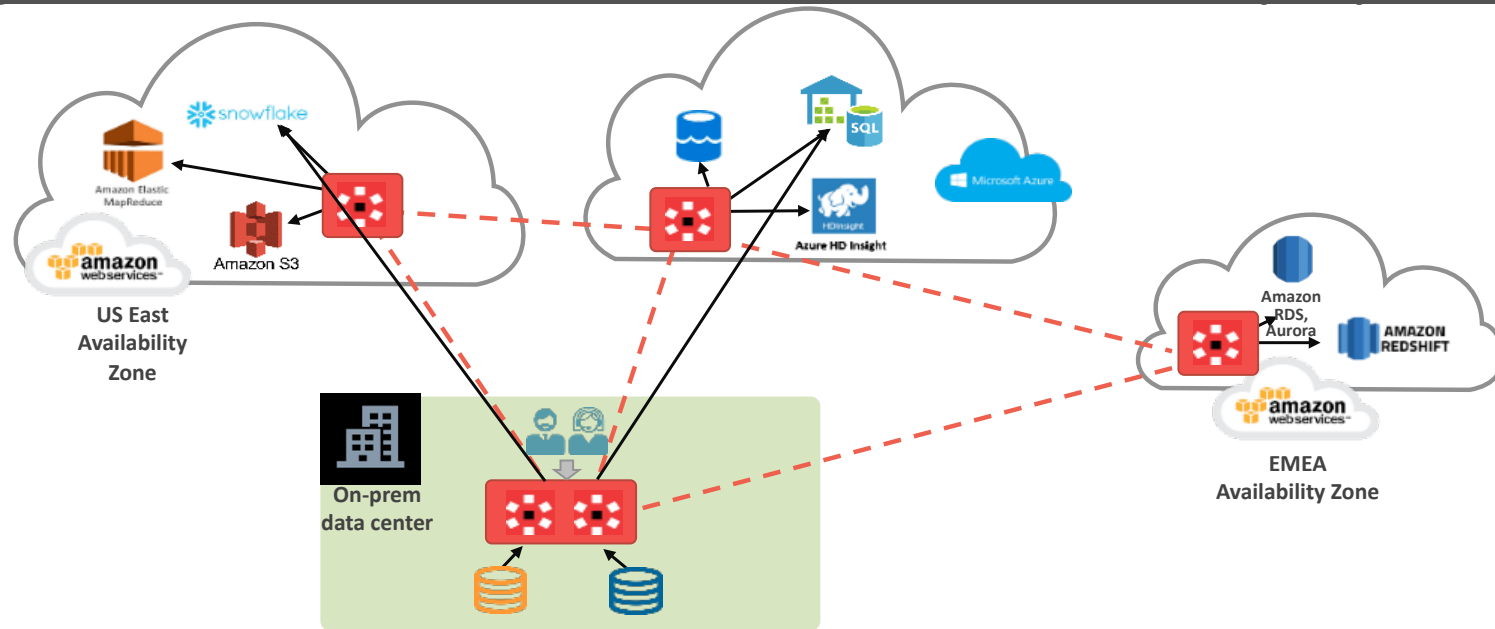
Acts as one Logical System – one semantic model
- User access any pt in system with Single Sign On

Servers are Distributed close to data sources
- Processing delegated to local sources
- Sending only result sets across network - minimizes data latency
- Caching and MPP to accelerate performance at each node

Logical Data Layer Advantages

1. **Decouples - Is Independent of:**
   - **Data Sources** - IT can use any data source w/o affecting Bus
   - **Applications & Tools** – Bus can use any tool w/o affecting IT
   - **Deployment** – multi-cloud, on-premise, hybrid, multi-geo, multi-platform

2. **Provides Real-time Enterprise Governance & Security**
   - Dynamically enforce security and access policies

3. **Uses ML/AI to Automate Tasks**
   - Optimizes Performance, Recommend best data sets to use
   - Implement 'Active Metadata' to enhance governance rules

denodo

Gartner Critical Capabilities for Data Integration Tools

Q3 2020

Score 1 -5, 5.0 being the highest.

Denodo is the only 5.0 of entire Ratings

Table 2. Product/Service Rating on Critical Capabilities

Critical Capabilities	Actian	Adeptia	Denodo	Fivetran	Hitachi Vantara	HVR	IBM	Informatica	Information Builders	Matillion	Microsoft	Oracle	Qlik	Safe Software	SAP	SAS	SnapLogic	Precisely	Talend	TIBCO Software
Bulk/Batch Data Movement	4.5	4.3	3.8	3.7	4.4	3.7	4.9	4.9	4.4	4.5	4.4	4.7	3.9	4.5	4.7	4.6	4.1	4.6	4.8	3.5
Data Virtualization	2.1	2.3	5.0	1.8	1.8	1.5	4.2	3.8	3.7	2.0	3.8	3.4	3.4	3.0	4.3	4.0	3.5	1.5	3.4	4.7
Message-Oriented Data Movement	3.5	4.1	3.9	1.0	3.0	3.9	4.0	3.9	4.0	1.0	4.0	3.8	3.6	3.3	4.0	3.2	4.0	3.0	4.0	4.5
Data Replication/Synchronization	3.9	2.8	2.5	4.5	2.8	5.0	4.4	4.6	3.9	4.0	4.0	5.0	5.0	4.3	4.5	4.0	4.0	4.2	4.3	3.8
Stream Data Integration	3.2	2.5	3.9	3.0	3.8	3.8	4.6	4.3	4.1	1.0	3.8	4.7	4.1	3.6	3.9	3.8	4.2	4.0	4.0	4.0
Data Preparation and Usability	3.8	4.3	4.4	3.0	4.1	2.0	3.7	4.1	2.9	3.0	4.4	3.6	3.6	4.4	3.5	4.7	3.5	3.6	4.3	3.0
Integration Style Interoperability	4.0	3.3	4.0	1.0	3.3	1.9	4.2	4.5	4.0	2.0	4.2	4.4	4.0	3.6	4.0	4.1	4.0	3.5	4.0	3.8
Active Metadata Support	1.6	3.2	3.6	2.6	3.5	2.5	4.4	4.6	3.6	2.0	3.5	3.8	3.9	3.8	4.0	3.9	3.9	2.6	3.8	3.0
Hybrid and Multicloud Support	3.5	4.0	4.3	3.9	3.5	4.0	4.4	4.5	2.4	3.8	3.0	3.3	3.9	4.1	3.6	4.2	4.5	3.5	4.4	3.8
Augmented Data Integration	1.0	2.5	3.4	4.2	3.0	2.2	4.5	4.7	2.2	2.6	4.4	4.6	1.0	3.3	4.1	4.1	4.1	3.6	3.9	3.6

As of June 2020

Source: Gartner (August 2020)