Data Virtualization: The rising alternative for agile data integration and data delivery

Suresh Chandrasekaran

Senior VP, Denodo

Agenda

1. Key Trends in IT – The need for Agile Data Delivery
2. Data Virtualization: An Introduction
3. 6 Solution Patterns – Where DV can be a Game Changer
4. Q&A

Key Trends in IT

Three Key Trends

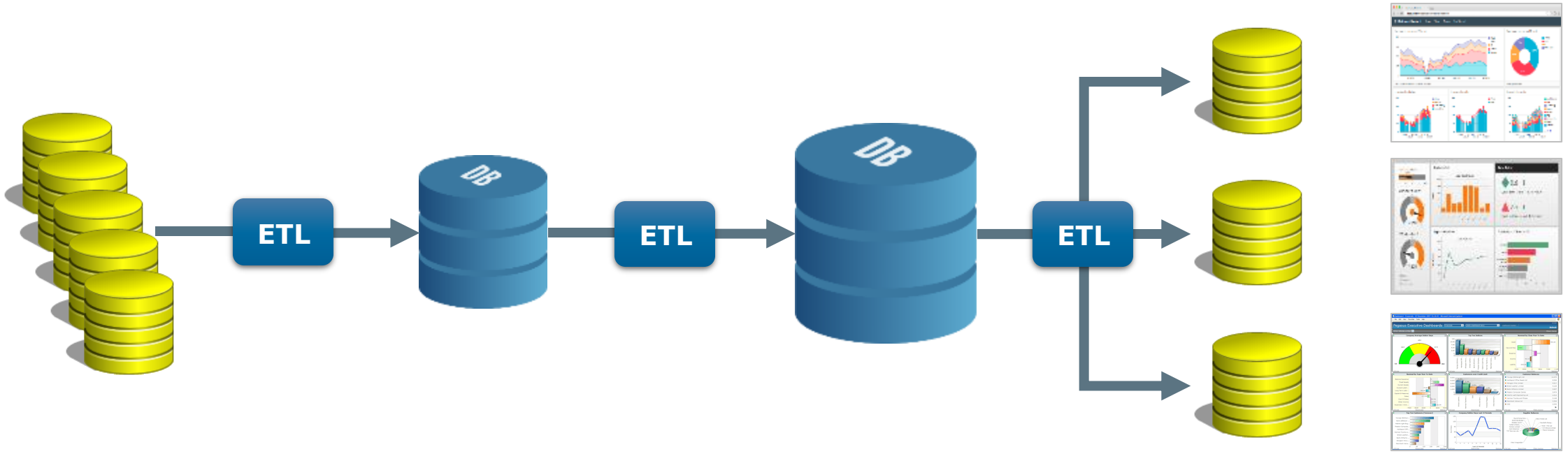
1. Reduce corporate data silos to gain efficiency and productivity
2. Towards a common data backbone for operational and informational use
3. Enterprises going with *bimodal IT* in their modernization efforts

What Does This Mean?

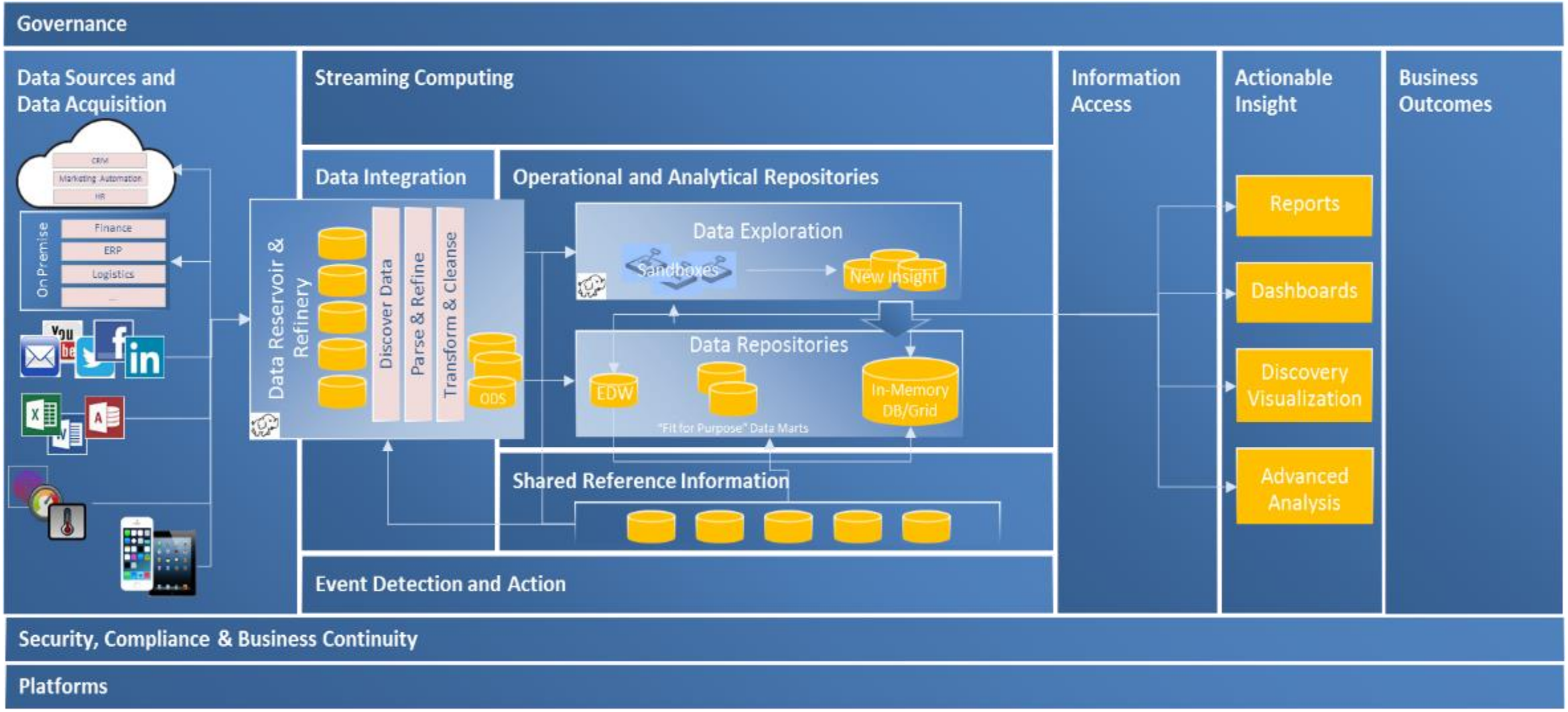
How does this affect the 'Information Architecture'?

- A data access layer is needed to 'open up' data silos
  - But retaining local ownership and control of the data
- The access layer must provide access to all data sources and support different modes of access
  - Reporting/analytics, real-time applications access (mobile/web and 'traditional'), etc.
- New technologies will be an important part of the information infrastructure
  - Hadoop ecosystem, NoSQL, streaming data, "Data Lakes"
- The traditional IT infrastructure is not going away soon
  - 'Systems of Record' still needed
- The new and the old need to work together
  - Newer systems still needs to interact with 'Systems of Record'

Data Integration – “The Way We Were…”



Data Integration – A Modern Data Ecosystem

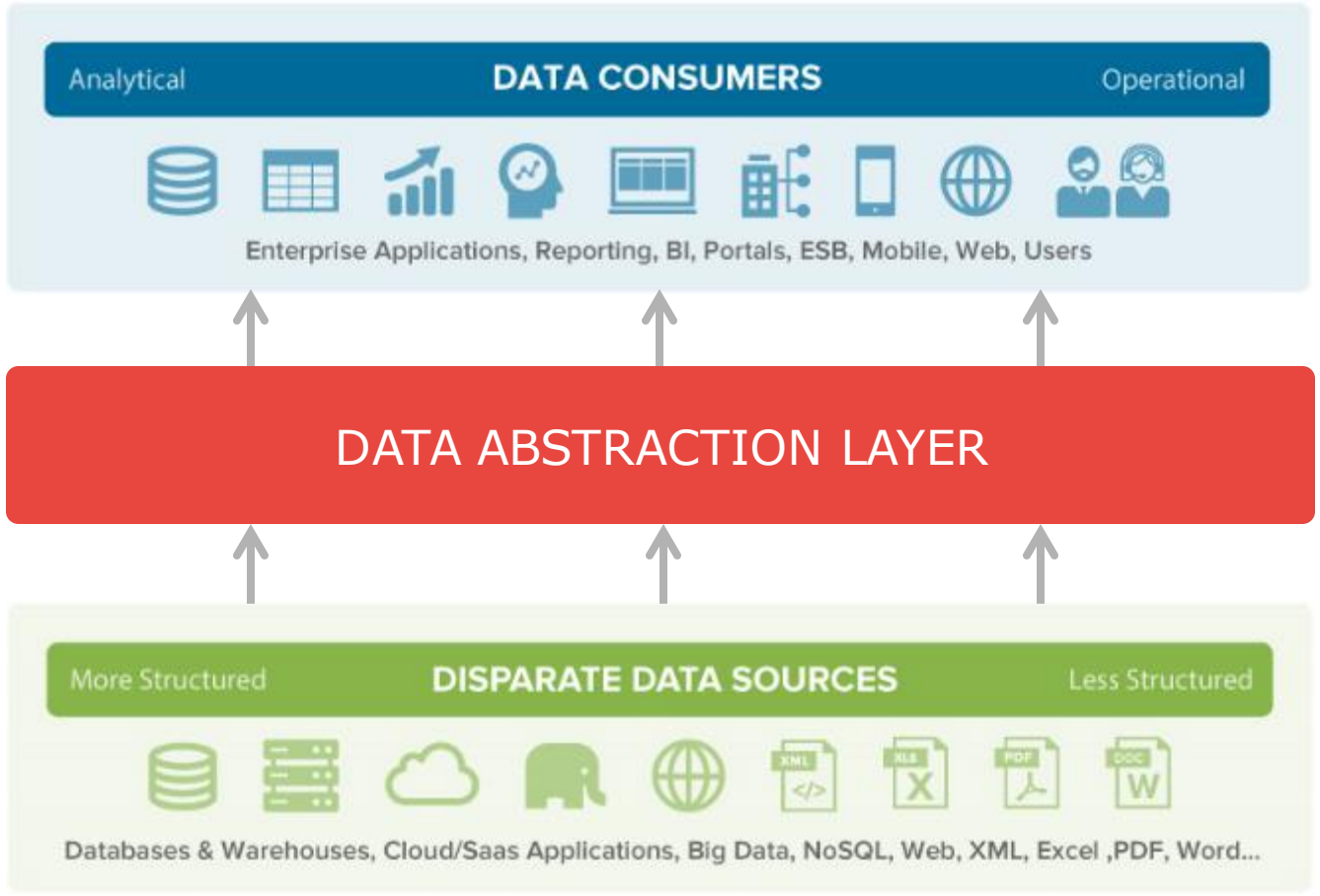


The Solution – A Data Abstraction Layer

- Abstracts access to **disparate data sources**
- Acts as a **single repository** (virtual)
- Makes data available in **real-time** to consumers

"Enterprise architects must revise their data architecture to meet the demand for fast data."

– Create a Road Map For A Real-time, Agile, Self-Service Data Platform, Forrester Research, Dec 16, 2015



Data Virtualization: An Introduction

Data Virtualization

Publishes
the data to applications

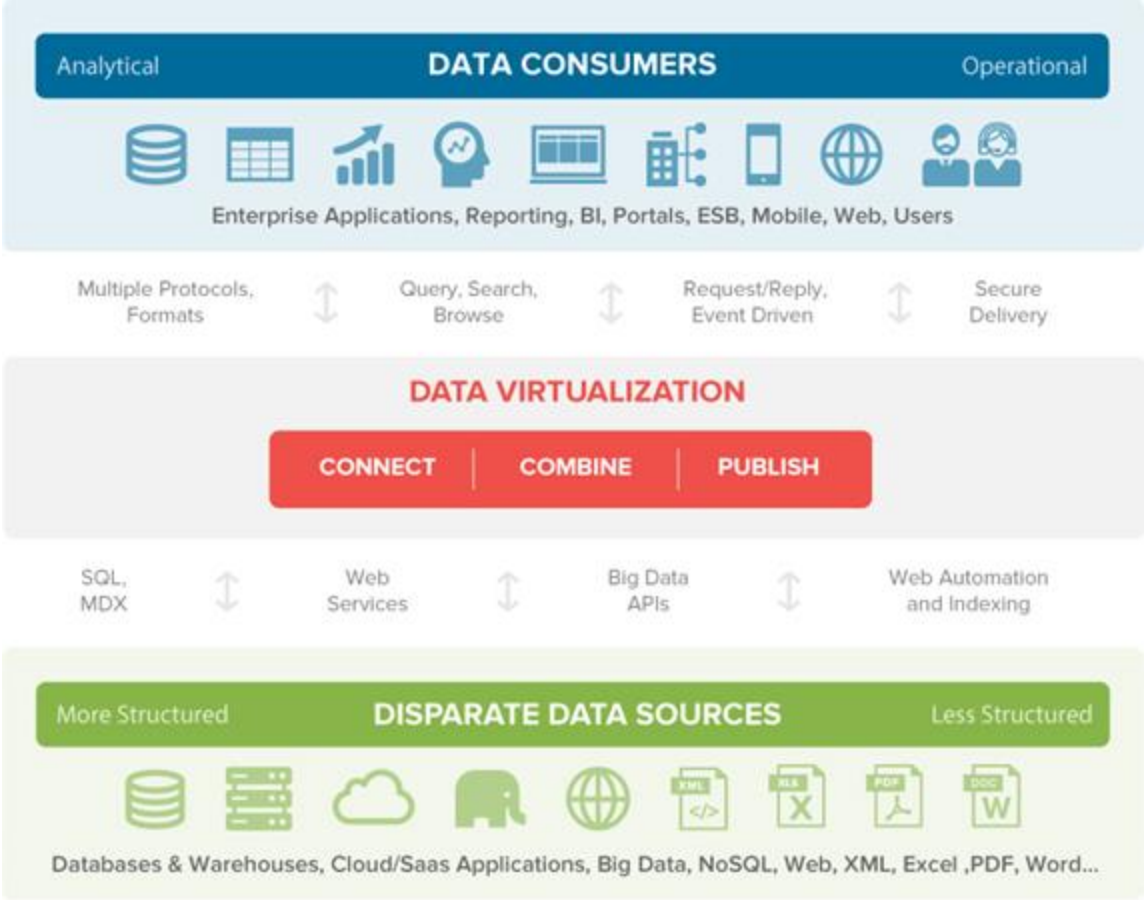
3

Combines
related data into views

2

Connects
to disparate data sources

1



"Data virtualization integrates disparate data sources in real time or near-real time to meet demands for analytics and transactional data."

– Create a Road Map For A Real-time, Agile, Self-Service Data Platform, Forrester Research, Dec 16, 2015

Data virtualization technology can be used to create virtualized and integrated views of data in memory (rather than executing data movement and physically storing integrated views in a target data structure), and provides a layer of abstraction above the physical implementation of data.

Source: "Gartner Market Guide for data virtualization – 2016"

What Data Virtualization is Not!

- It is not ETL
  - If you want to replicate data from 'A' to 'B'…use an ETL tool – it's what they are designed for
- It is not Data Visualization (← Note the 's')
  - It complements visualization and reporting tools (e.g. Tableau)
- It is not a database
  - Data Virtualization Platforms don't store the data…it's retrieved from the data sources on demand
- It has many capabilities such as governance, metadata management, security, etc.
  - It will work with specialized tools in these areas
- It's great for service-based architectures
  - But be wary of event-driven architectures…use an ESB (or similar) for this

Data Virtualization Platforms – Key Capabilities

Five Essential Capabilities of Data Virtualization

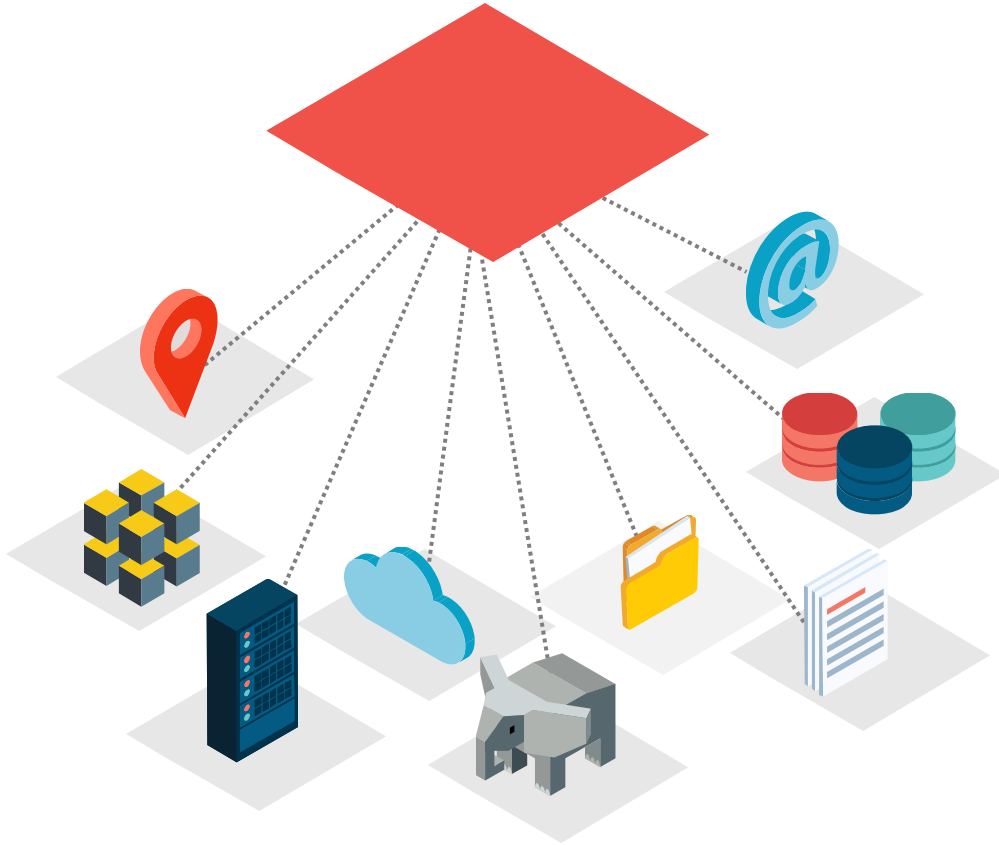
1. Data abstraction
2. Zero replication, zero relocation
3. Real-time information

4. Self-service data services
5. Centralized metadata, security & governance

1. Data abstraction

…hides data complexity for ease of data access by business

- Abstracts access to **disparate data sources.**
- Acts as a **single virtual repository.**
- Abstracts data complexities like **location, format, protocols**

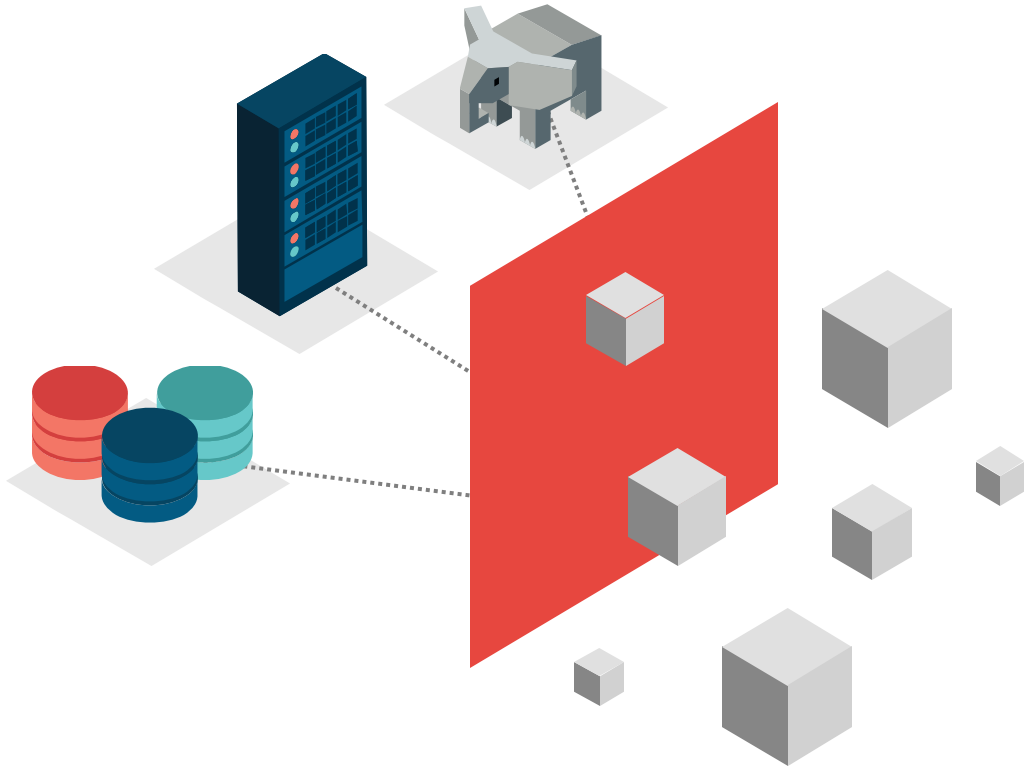


> **Enterprise architects must revise their data architecture to meet the demand for fast data.”**
>
> *– Create a Road Map For A Real-time, Agile, Self-Service Data Platform, Forrester Research*

2. Zero replication, zero relocation

…reduces development time and overall TCO

- Leaves the data at its source; **extracts only what is needed**, on demand.
- Diminishes the need for effort-intensive ETL processes.
- **Eliminates unnecessary data redundancy**.



“The Denodo Platform enables us to build and deliver data services, to our internal and external consumers, within a day instead of the 1 – 2 weeks it would take with ETL.”

– Manager, DrillingInfo

3. Real-time information

…enables timely decision-making

- Provisions data in **real-time** to consumers
- Creates real-time **logical views of data** across many data sources.
- **Supports transformations and quality functions** without the latency, redundancy, and rigidity of legacy approaches

> **Data virtualization integrates disparate data sources in real time or near-real time to meet demands for analytics and transactional data."**

– Create a Road Map For A Real-time, Agile, Self-Service Data Platform, Forrester Research, Dec 16, 2015

3. Real-time integrated information

…enables timely decision-making

- Dynamic cost-based query optimization
  - Designed for federated queries
- Data source constraint optimized
- Advanced optimizations for very large datasets
  - e.g. Partial aggregation delegation
- Intelligent caching
  - Selective materialization
- Transformations, integration performed ‘on the fly’
- Minimal overhead

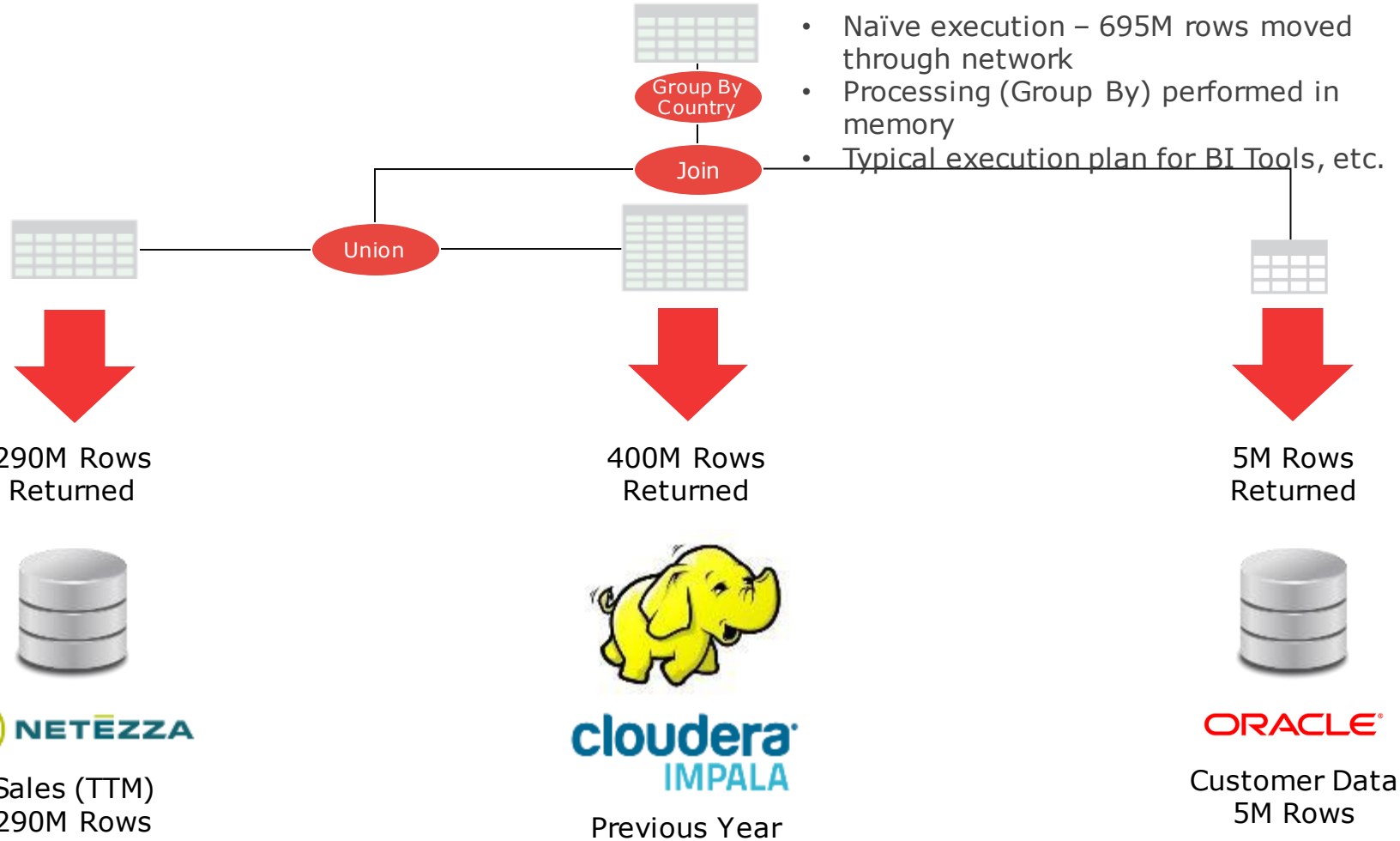
Execution Strategy	Execution Time
Naïve Execution Strategy – BI Tools, Other Data Federation	> 20 mins
Denodo 6.0 – Partial aggregation pushdown	51 secs
Denodo 7.0 – Partial aggregation pushdown and integrated in-memory fabric	24 secs

Dataset: TPC-DS dataset
Query: Total Sales by Customer Country in Last Two Years
Data Sources: Netezza (Sales TTM – 290M rows)
Hadoop (Sales Past 5 Years – 2.9B rows)
Oracle (Customers – 5M rows)

3. Performance Acceleration – Naïve Approach

> 20 mi

Query: Total Sales by Customer Country in Last Two Years

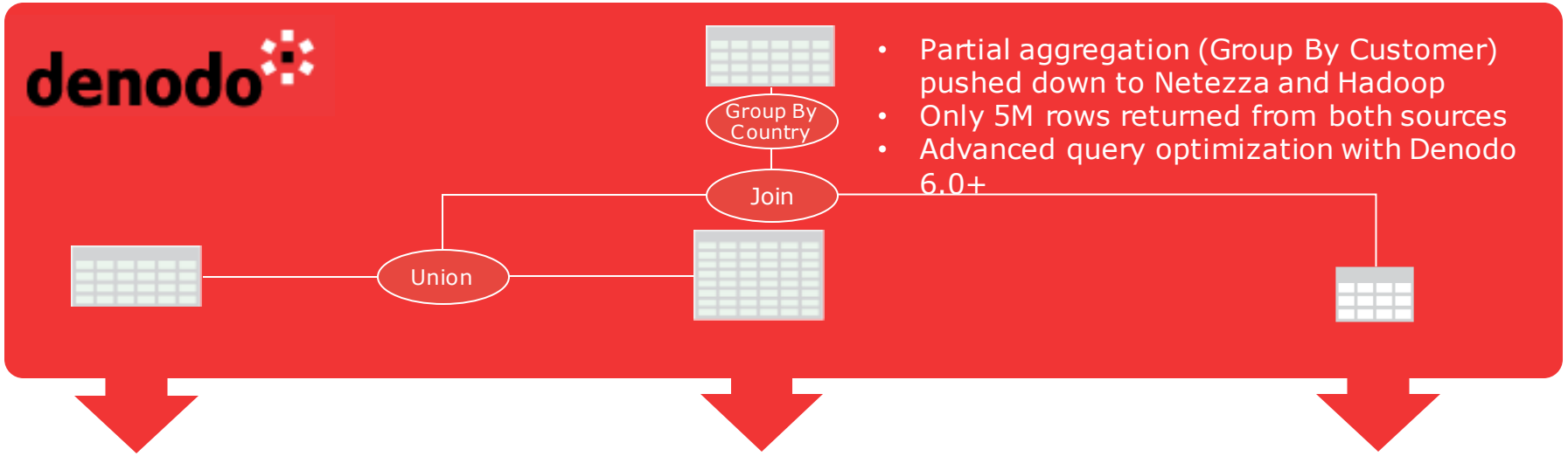


- Naïve execution – 695M rows moved through network
- Processing (Group By) performed in memory
- Typical execution plan for BI Tools, etc.

3. Performance Acceleration – Denodo 6.0

51 se

Query: Total Sales by Customer Country in Last Two Years



5M Rows Returned

Group By Customer

NETEZZA

Sales (TTM)
290M Rows

5M Rows Returned

Group By Customer

cloudera IMPALA

Previous Year Sales
400M Rows

5M Rows Returned

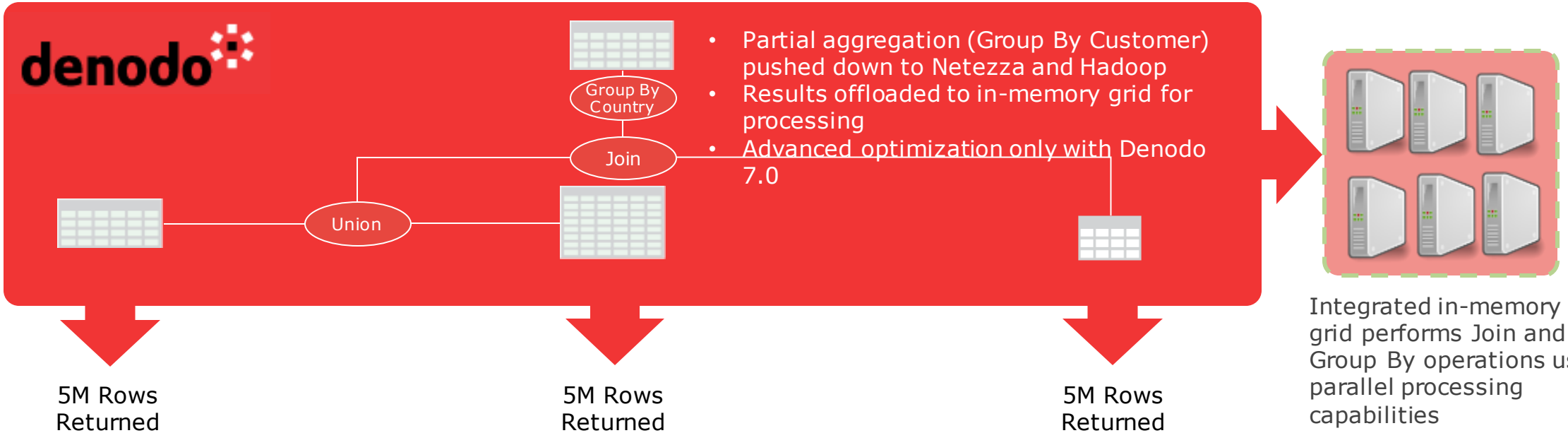
ORACLE

Customer Data
5M Rows

3. Performance Acceleration – Denodo 7.0

24 se

Query: Total Sales by Customer Country in Last Two Years



4. Self-service data services

…enables information discovery and self-service

- Facilitates access to all data, both internal and external
- Enables creation of **universal semantic models** reflecting business taxonomy
- Connects data silos to provide best available information to drive business decisions

> **Impressively quick turn around time to "unlock" data from additional siloes and from legacy systems - Few vendors (if any) can compete with Denodo's support of the Restful /OData standard - both to provide data (northbound) and to access data from the sources (southbound).”**
>
> *– Business Analyst, Swiss Re*

5. Centralized metadata, security & governance

…simplifies data security, privacy, audit

- Abstracts data source security models and enables single-point security and governance.
- Extends single-point control across cloud and on-premises architectures
- Provides multiple forms of metadata (technical, business, operational) to facilitate understanding of data.

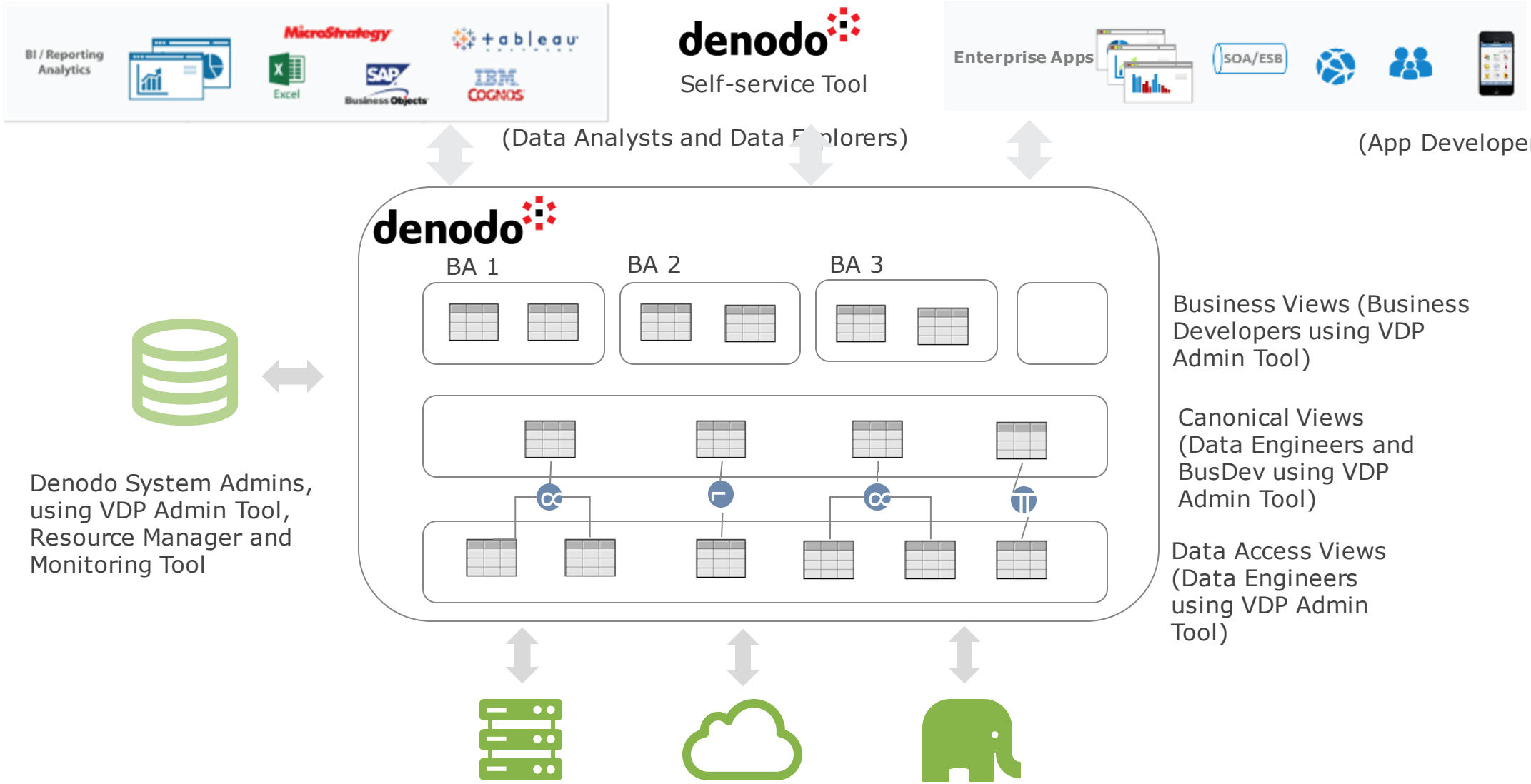
> **Our Denodo rollout was one of the easiest and most successful rollouts of critical enterprise software I have seen. It was successful in handling our initial, security, use case immediately, and has since shown a strong ability to cover additional use cases, in particular acting as a Data Abstraction Layer via it's web service functionality."**
>
> *– Enterprise Architect, Asurion*

Information Self-Service & Data Catalog

Extended metadata and active catalog of enterprise data assets

- **Catalog** business views by server/folder, business categories and tags
- New Roles and permissions to manage the catalog
- **New UI** and improved graphical transformations

Information Self-Service Architecture with Denodo



Data Virtualization: Solutions

Denodo Adds Value to Entire CxO Agenda

Data-Centricity in Business

Innovating through big data, adding new sources for enterprise use, Advanced Analytics

Focus:

Integration layer for Big Data & other corporate data assets. IoT innovative projects. Data-driven business models. Data science and analytics.

Enabling Business Agility

The Enterprise Data Marketplace. Enabling Self-Service

Focus:

Making LOB partners agile i.e. launch new products, get closer to customer, offer data visibility and rapid data provisioning

Efficiency in Data Operations

Driving operational efficiencies and reduced cost

Focus:

Reduce costs and complexity, minimize data replication, foster data reusability and collaboration

Taming the data "mess"

Unifying a diverse universe of data assets and helping to enforce enterprise data policies

Focus:

Data Governance, Discovery, Unified Data Modeling, Security, Data Auditing

Denodo 'Solution' Categories

Customer Centricity / MDM

- ✓ Complete View of Customer

Data Governance

- ✓ GRC
- ✓ GDPR
- ✓ Data Privacy / Masking

Data Services

- ✓ Data as a Service
- ✓ Data Marketplace
- ✓ Data Services
- ✓ Application and Data Migration

BI and Analytics

- ✓ Self-Service Analytics
- ✓ Logical Data Warehouse
- ✓ Enterprise Data Fabric

Big Data

- ✓ Logical Data Lake
- ✓ Data Warehouse Offloading
- ✓ IoT Analytics

Cloud Solutions

- ✓ Cloud Modernization
- ✓ Cloud Analytics
- ✓ Hybrid Data Fabric

Intel: POC in 2013 - "Game Changer" in 2015

Solutions that deploy quickly and provide a performance boost

Download PDF

IT@Intel White Paper

Integrating Data Warehouses with Data Virtualization for BI Agility

Executive Overview

RESULTS

Data virtualization proved an agile solution with both of our tools, requiring only one week to set up compared to the typical eight weeks it takes to implement a co-location solution. In performance, data virtualization posted better times than single source co-location in scenarios where we merged data from two data sources and used filters to push down processing to the two data sources.

Compared to the time it takes to model, design, develop, and test a co-location setup, data virtualization is much easier and less expensive to set up, test, and put in operation. This makes it a promising solution for situations where a business group needs a fast reporting solution. In addition, once a data virtualization solution is in place, adding a connection to a new source in the virtual layer is quick and easy.

CONCLUSION

While only certain scenarios in our PoC benefited from our current data virtualization solutions, the growing importance of implementing agile solutions for extracting business intelligence from big data and traditional enterprise data suggests that such methods will grow in importance and sophistication. Our findings provide initial guidance for the use of data virtualization at Intel and are driving our interest in investigating dedicated data virtualization solutions.

Based on our PoC results, when faced with a need for data integration across multiple data containers, we will now try data virtualization solutions, using rapid prototyping to see if virtualization can provide business value. If it can, we will expand its use accordingly. If we experience performance bottlenecks, complex transformations, or other issues that impact the desired results, we will switch to a co-location solution.

Data Virtualization is a "Game Changer" for Intel's Data

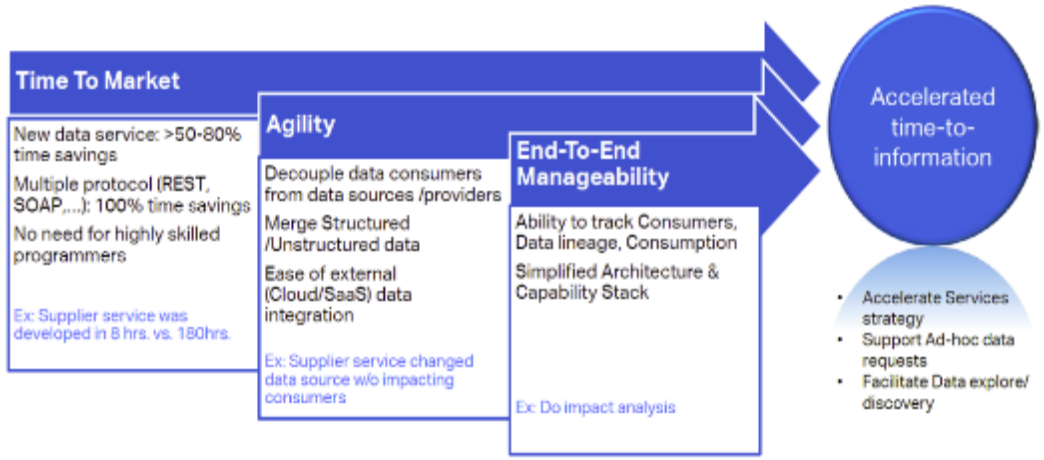
Guillermo Rueda, Atul Borkar & Stacie Hall

Enterprise Data World 2015, Washington, DC

March 29 – April 3, 2015

Intel Information Technology

DV Benefits



Intel: DV Benefits, Detail Metrics

Value Driver	Metric	Goal	Actual
Time to Develop	Time to develop web service in days	50%	90%
Time to Deploy	Time to Deploy web service in days	50%	90%
TTM	Overall time it takes to make web service available for use	60%	90%
Time to Engage	Time it takes for business to engage with IT	75%	75%
Performance	Performance of web services	50%	60%
Impact Analysis	How fast can we perform impact analysis	50%	90%
Enterprise Architectural Alignment	Ease at which data from disparate sources can be integrated	Security, data classification	High

Intel: Rapid Enterprise-wide Deployment

- 2013 – 10 DV trained staff
- 2016 – 650+ DV trained staff

- 2013 – Single project team
- 2016 – Intel DV CoE guiding 16/26 BU's in DV Project Use

- 2013 – Initial purchase for HR project
- 2016 – 3 year ELA – extended for another 3; many projects

- 2013 – <10 data sources, single server
- 2016 – 309 data sources, 100's of clustered servers (with HA and DR)

Intel’s Data Virtualization Capability

Central Tenets

- Intel is focused on reuse
- Time to market is a key value proposition
- Training is required and governance is enforced

Key Learnings

- Intel IT culture is not naturally suited for reuse.
- Security, as with any middleware, must be designed in and factored for.
- Clear guidance around DV vs other capabilities must be agreed upon and communicated

DV at Intel

- 650 internal users trained on DV tools and capability in 2x4-hour in-house built training session.
- 16/26 logical business areas now using data virtualization.
- 1263 distinct data services in the path to production from 309 data sources.
- All IT targets for business adoption have been exceeded.

Standards

- Enterprise data model aligned
- SOA and ETL compliance
- Light weight governance
- API registered services

Service

- Certified ITS data handling
- Robust, HA, DR infrastructure
- Rapid fix/feature deployments
- Fast and easy connectivity

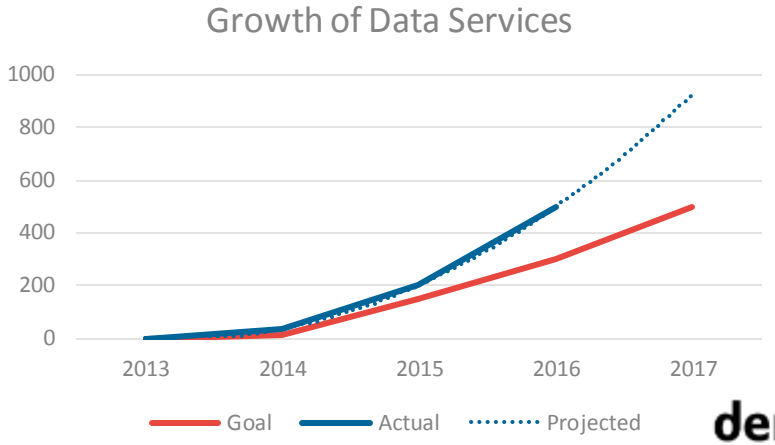
Support

- 24x7 SE support
- CE project design support
- Training, PB, Wiki, Videos
- Monthly COP meet-ups

Influence for Growth

Create Platform Engagement at All Levels

- Executive Messaging
  Create critical success indicators, measure progress to plan, communicate milestones
- Customer Messaging
  Innovate your platform and capabilities, communicate your wins and showcase your customers
- Vendor Messaging
  Influence through regular engagements, request platform enhancements, report internally and externally on vendor support



denodo

Database name
servicemanagement

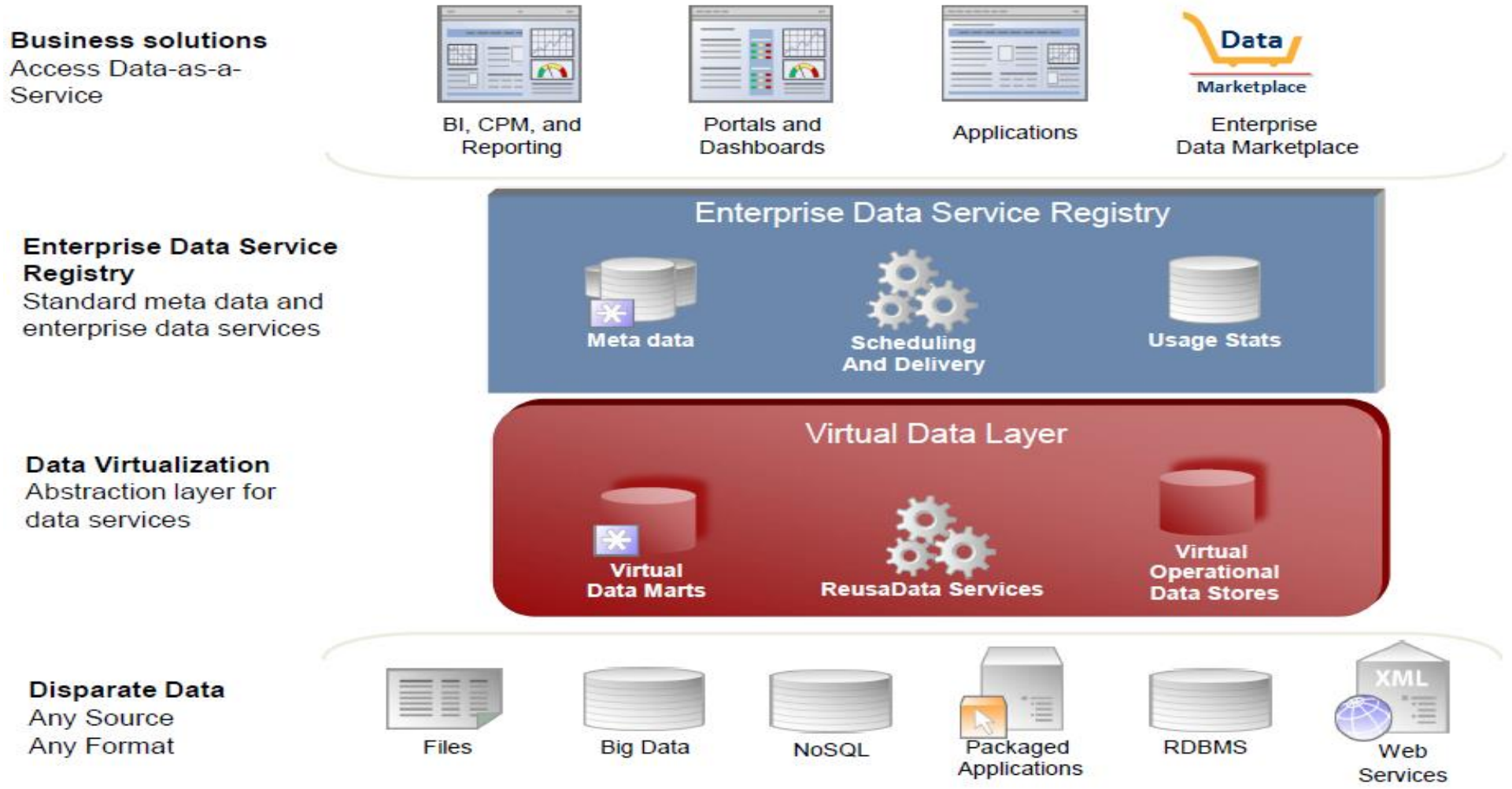
Description
Enterprise Data Virtualization Platform

IntelObjectType	Count
Distinct Data Services	1263
Datasource	309
Business View	1705
Base View	866
Integrated View	1216
Web Service	403

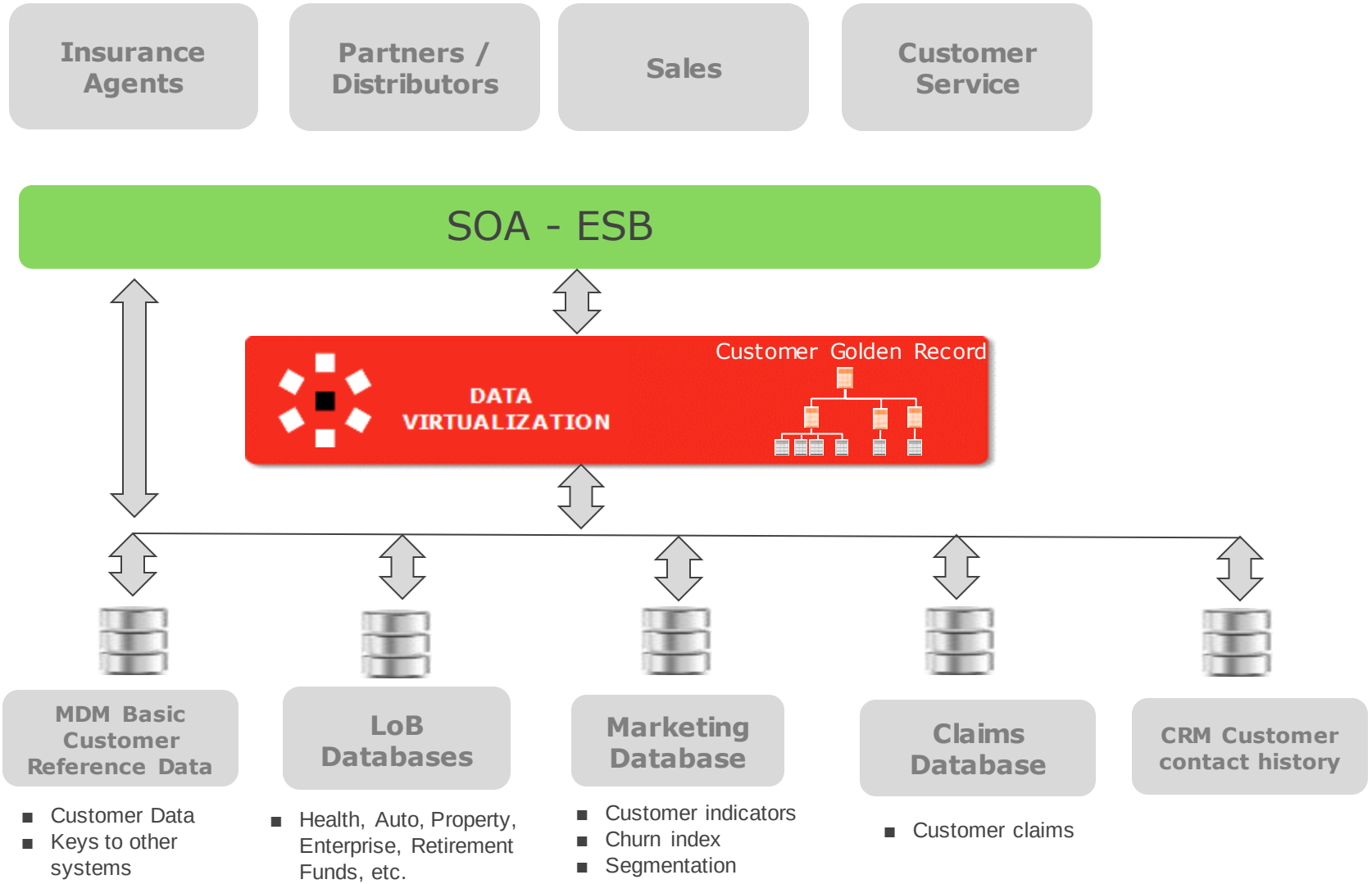
MY ENHANCEMENTS

Check your enhancement requests along with their status

Guardian Life Data Services Marketplace



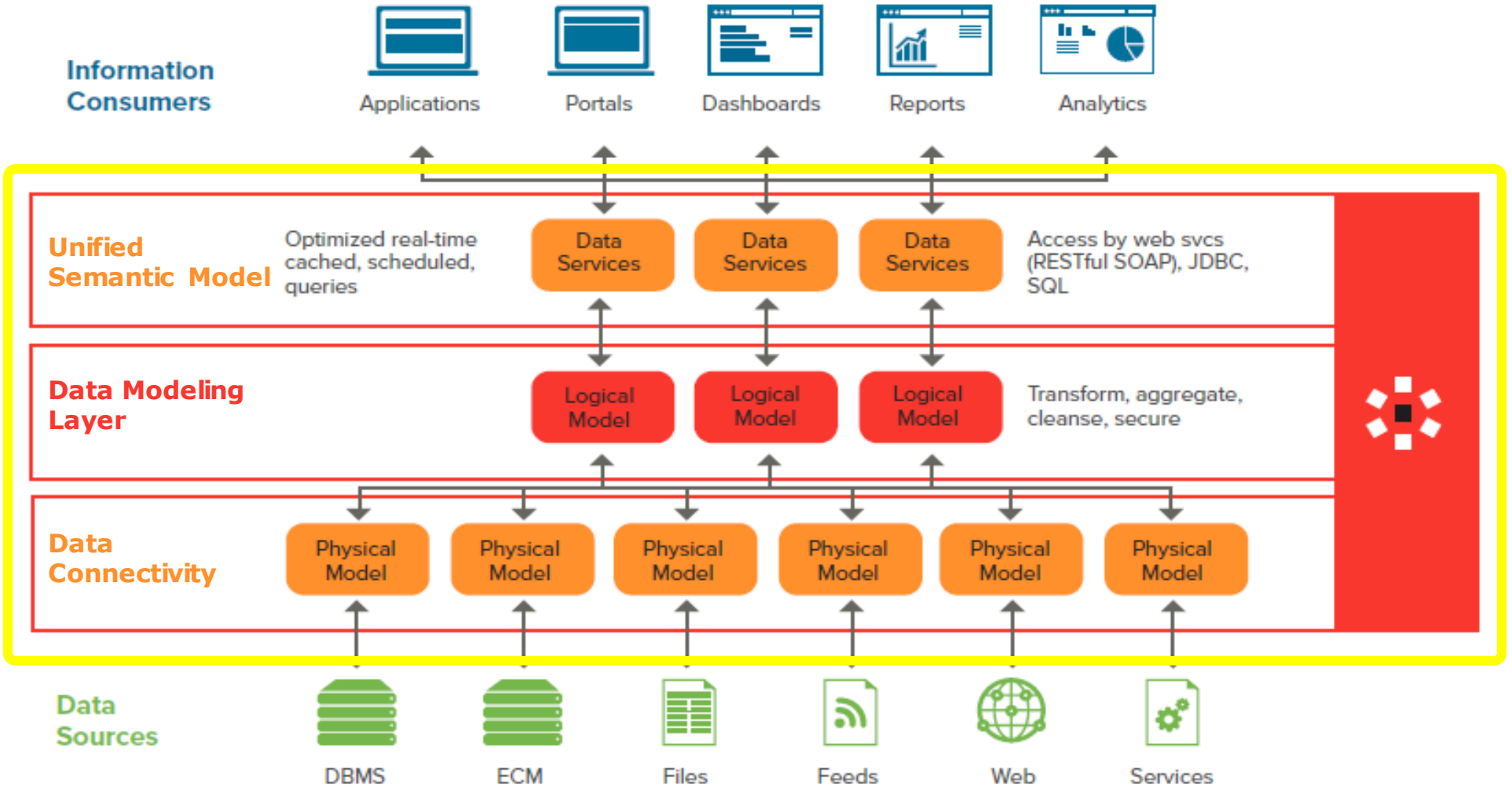
Unified Desktop for Complete View of Customer



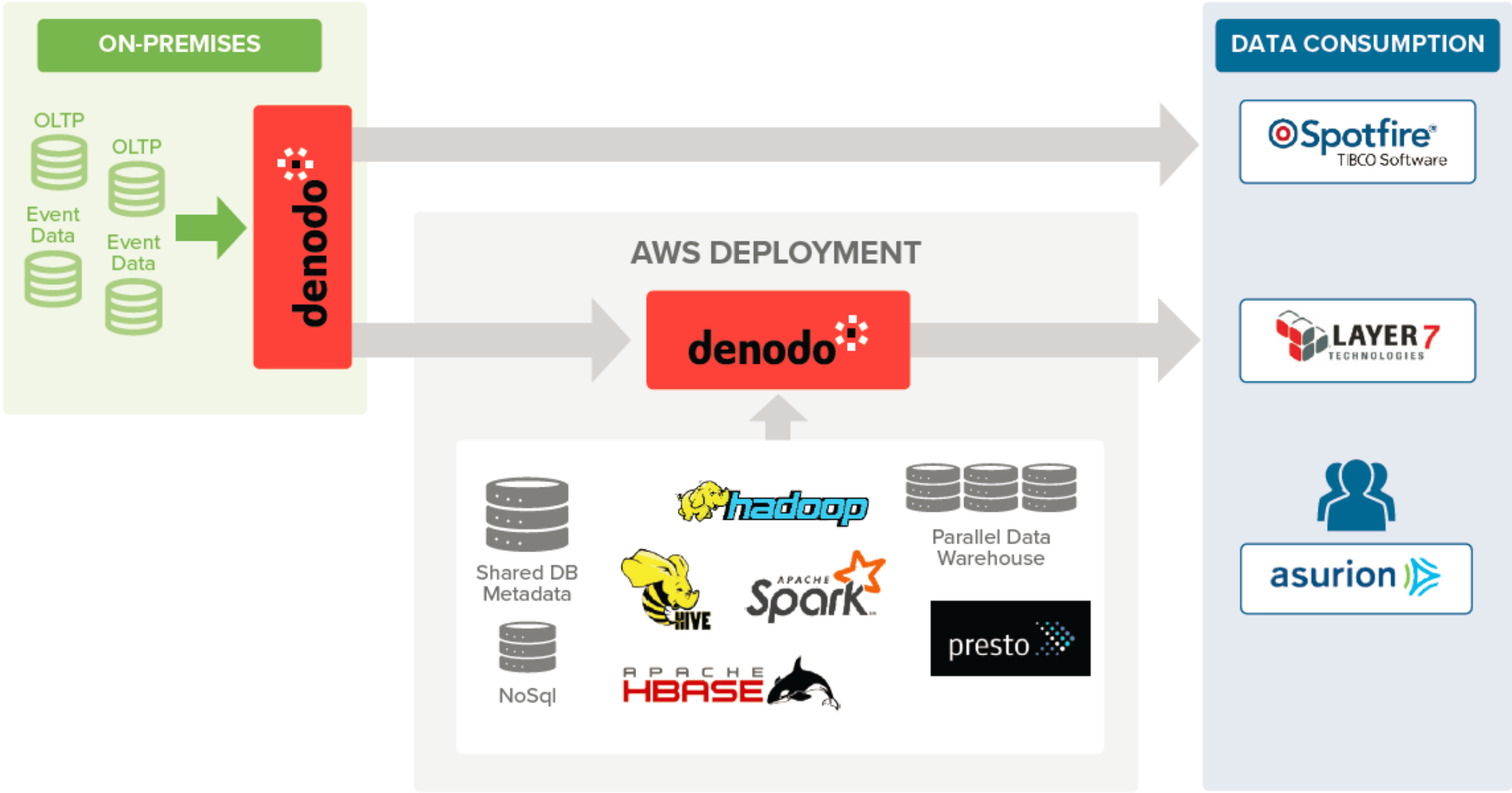
Benefits

- Unified Data Services Layer
- Integrates MDM basic data with transactional data
- Provides data services to multiple consuming applications: agents, distributors, sales, customer service,
- Single Customer Source of Truth across business lines
- Integrated with ESB through Web Services

Customer Architecture

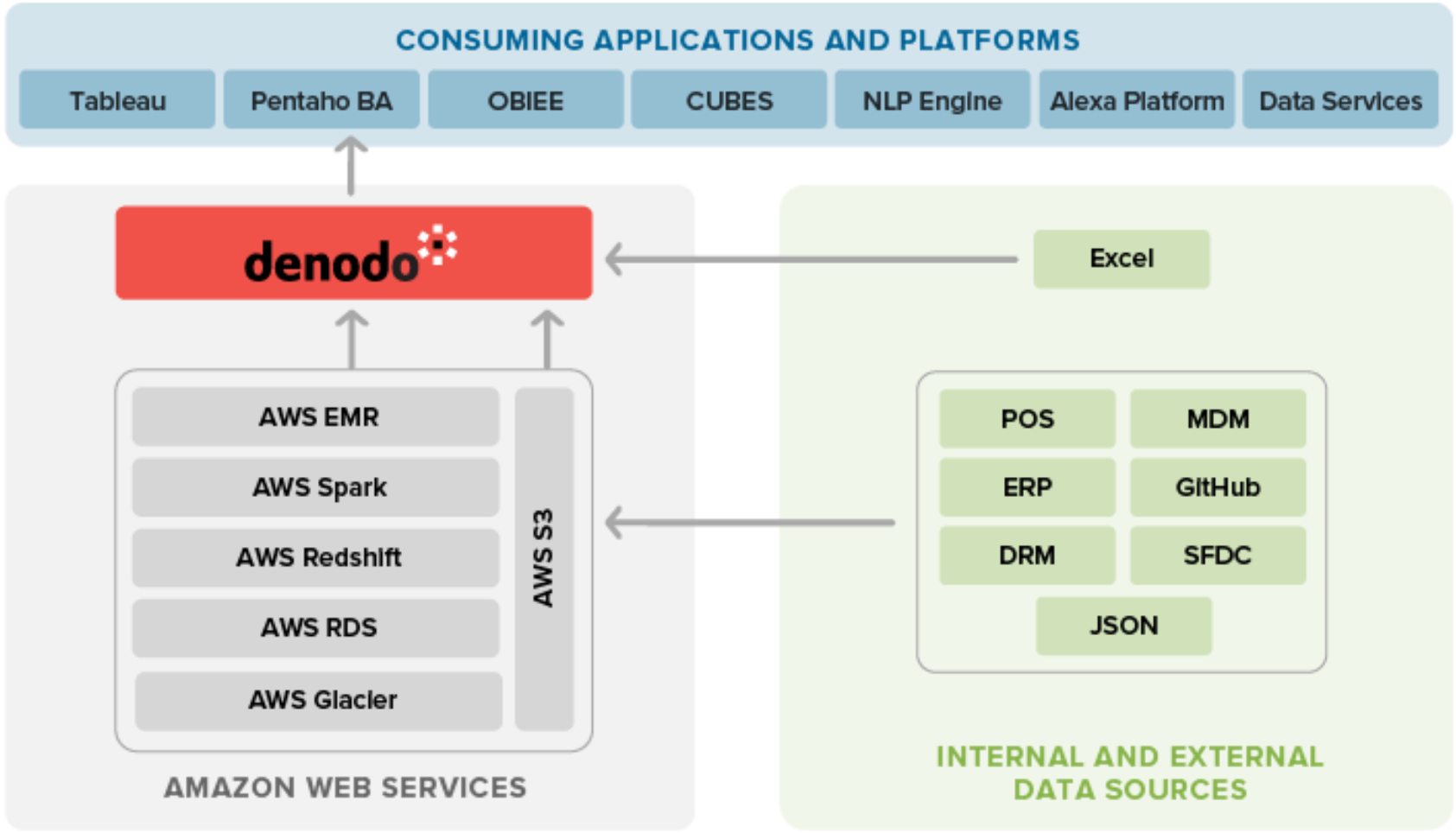


Asurion – Hybrid Architecture



Logitech

Architectural Diagrams

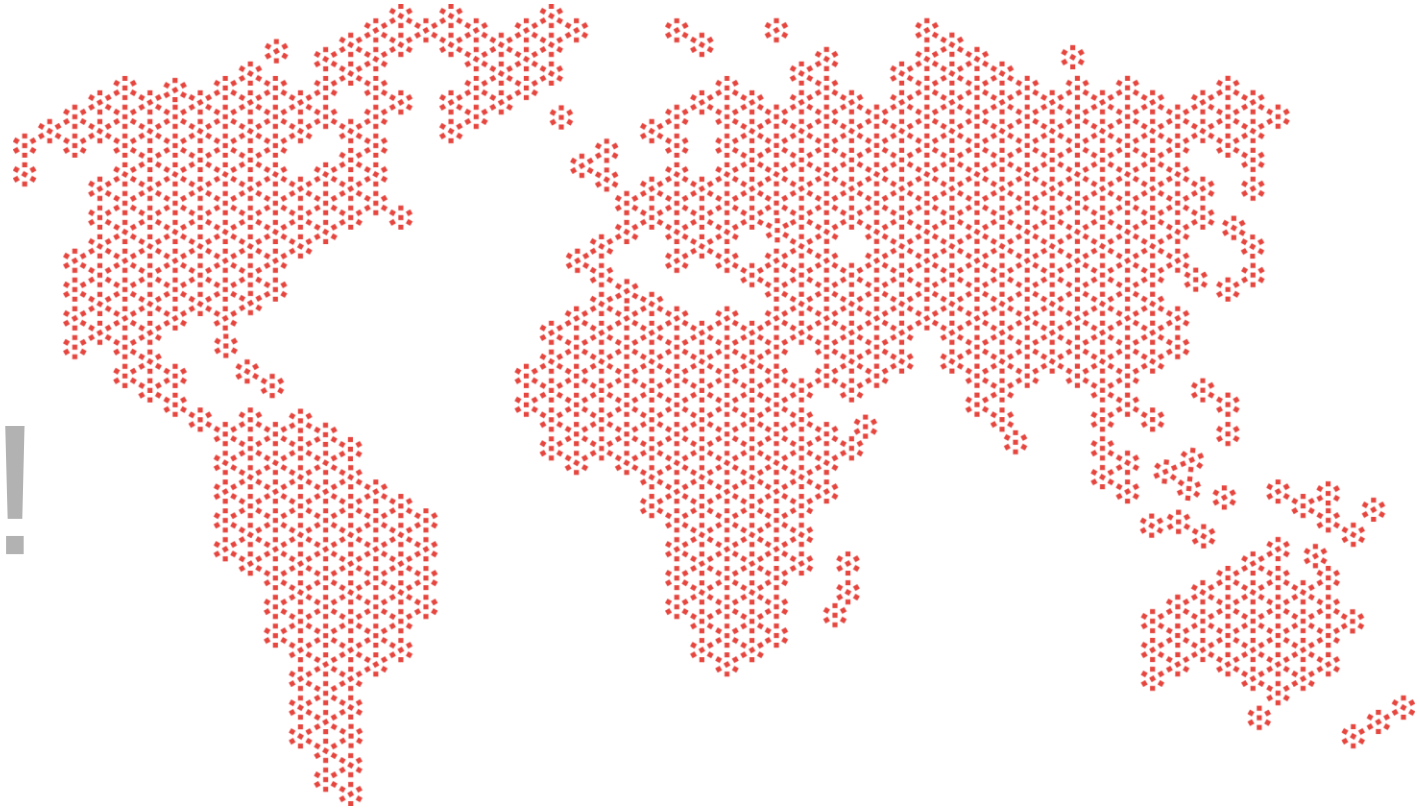


Summary

- Data Virtualization offers an alternative approach to data integration and provisioning
  - Very agile, iterative approach – involve business stakeholders at beginning and deliver quickly
  - Does not use expensive bulk data loads for data consolidation and centralization
  - Data is delivered on-demand, in real-time
- Five key capabilities of a Data Virtualization Platform
  - Data abstraction, Zero (minimal) replication, Real-time information, Self-service information, Centralized metadata, security, and governance
- Wide range of use cases and solution patterns
  - From Analytics and Big Data to Cloud to Governance and Data Services
- Data Virtualization – a "game changer" for agile data delivery and access

Q&A

Thanks!



denodo | www.denodo.com | info@denodo.com

© Copyright Denodo Technologies. All rights reserved
Unless otherwise specified, no part of this PDF file may be reproduced or utilized in any for or by any means, electronic or mechanical, including photocopying and microfilm, without prior the written authorization from Denodo Technologies.