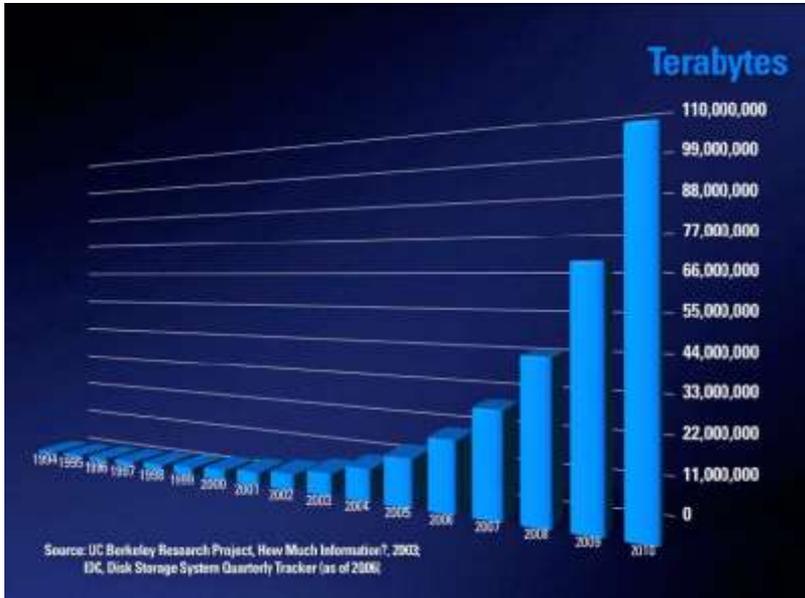




*Data Virtualization and  
Information Sharing in  
the Federal Environment*

David Besemer  
CTO, Composite Software

# Data Volume and Complexity are Exploding

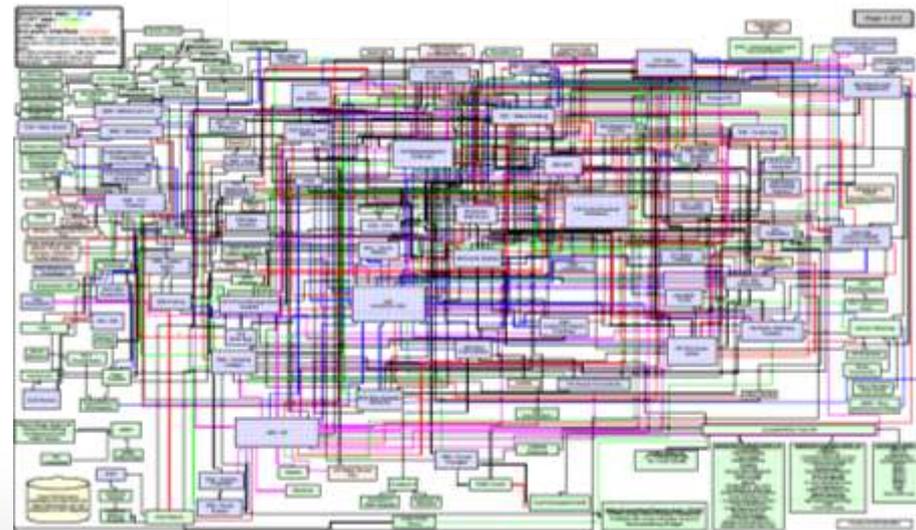


*“Data is growing is growing by a **factor of 10 every five years**, a compound annual **growth rate of almost 60%.**”*

*Source: IDC*

*“Managers spend **two hours a day** looking for information they need, and almost **half the data is useless** once they get it”*

*Source: Accenture*

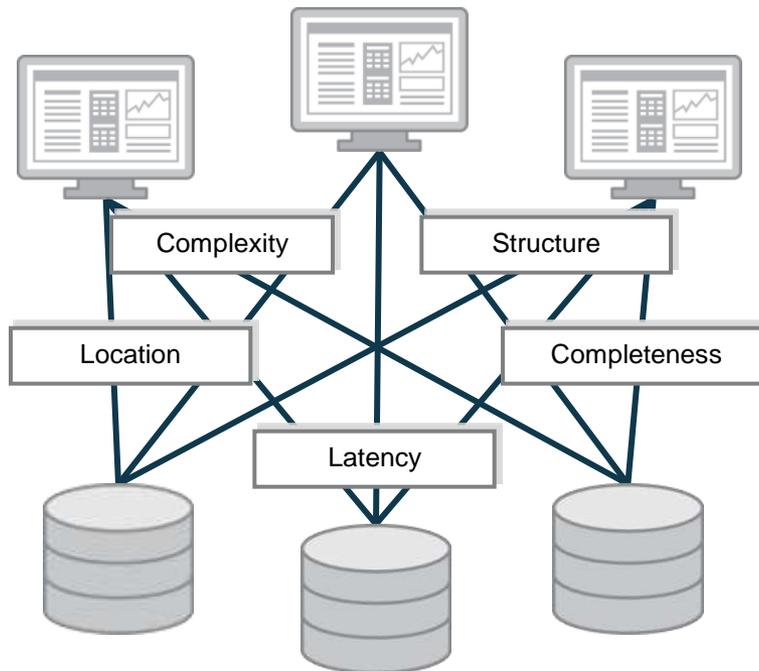


# The Challenge Data Virtualization Addresses

**Agency Objectives**



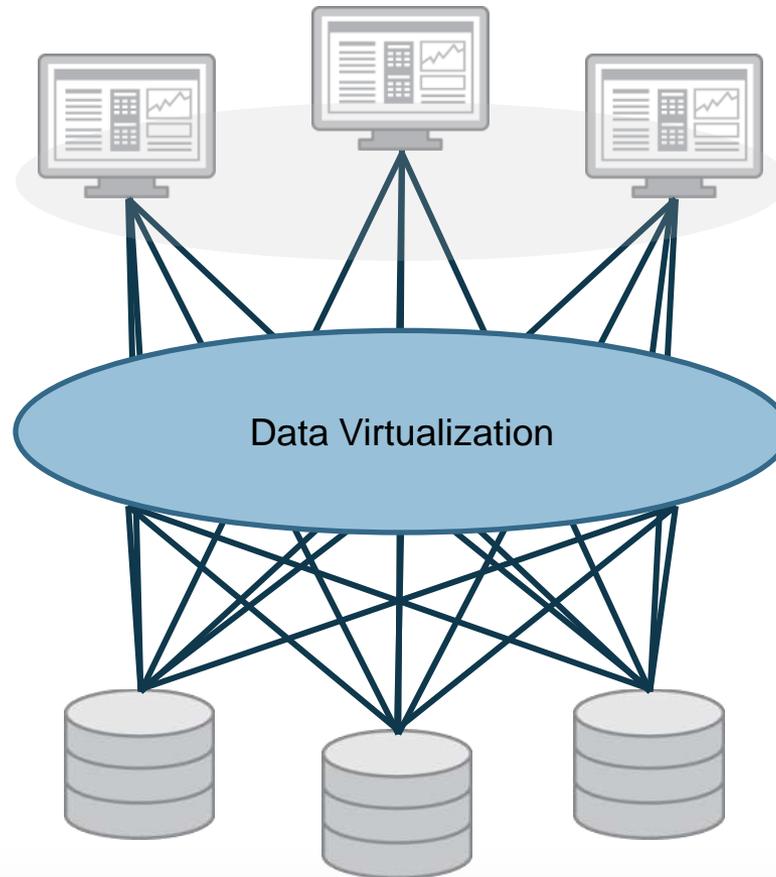
**Agency Solutions**



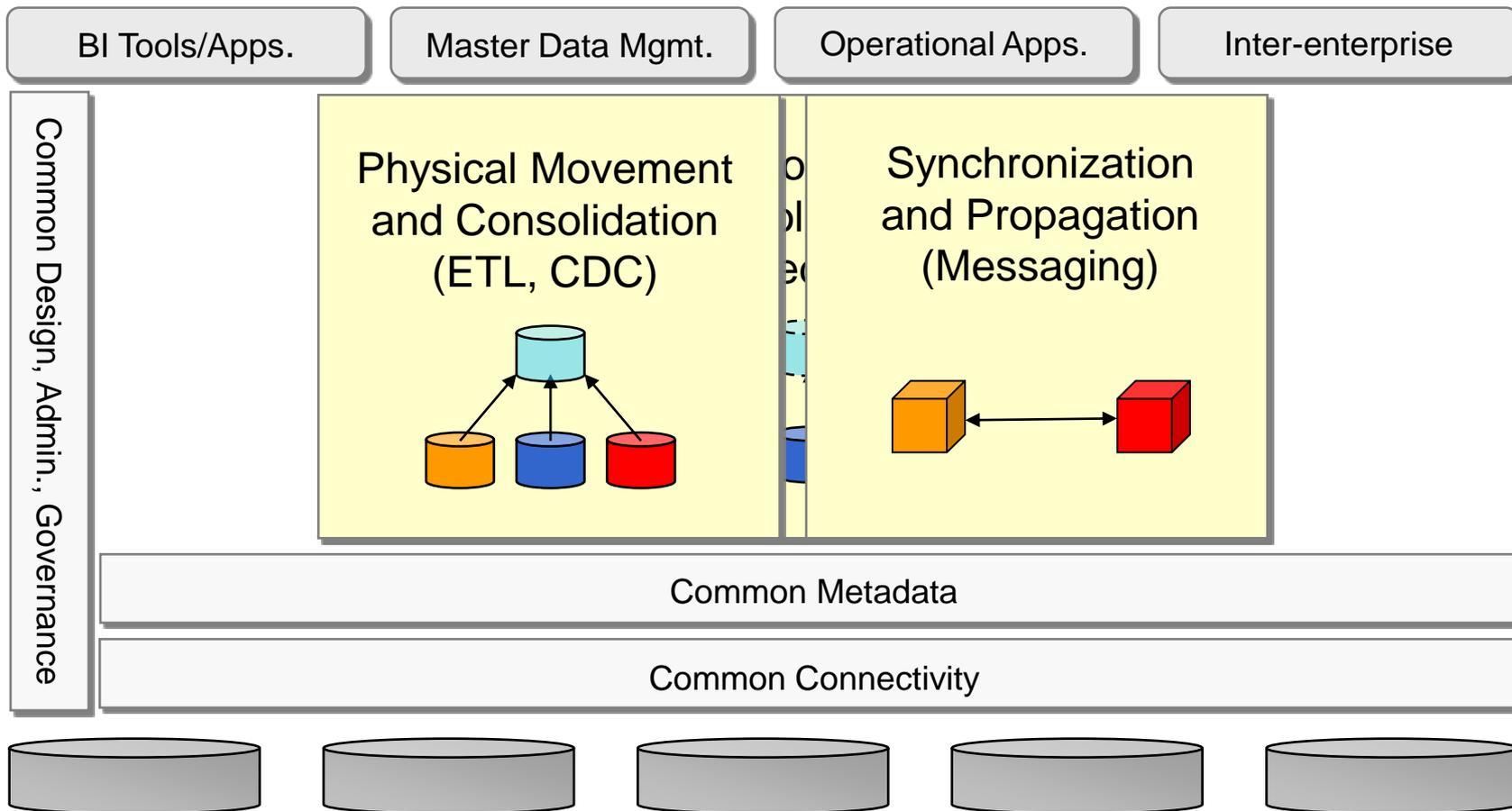
**Data Integration Challenge**

**Data Sources**

# What Data Virtualization Does



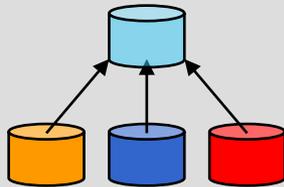
# Data Integration Architectures and Patterns: Build a Portfolio to Address the Range of Needs



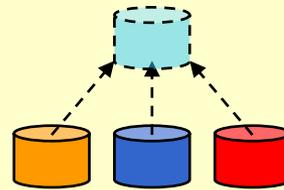
**Leading organizations support multiple styles of data integration and delivery to address a range of business requirements — breadth enables leverage and agility.**

# How Data Virtualization Differs

Physical Movement  
and Consolidation  
(ETL, CDC)



Abstraction / Virtual  
Consolidation  
(Data Federation)



Synchronization  
and Propagation  
(Messaging)



Middleware	ETL	CDC	<i>Data Virtualization</i>	EAI / ESB
Purpose	DB ↔ DB	DB ↔ DB	<i>DB ↔ Application</i>	Application ↔ Application
Attribute	Scheduled	Event Driven	<i>On Demand</i>	Event Driven

INFORMATICA GOLDENGATE

COMPOSITE  
— SOFTWARE —

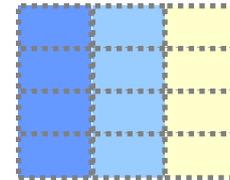
TIBCO  
The Power of Now®

# How Data Virtualization Works – Example Scenario

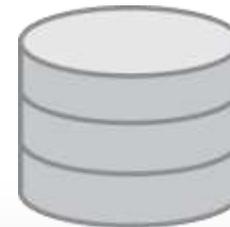
**1) I need to build an application that looks like this...**



**2) The view or data service needs to look like this...**



**3) And the data comes from these sources...**



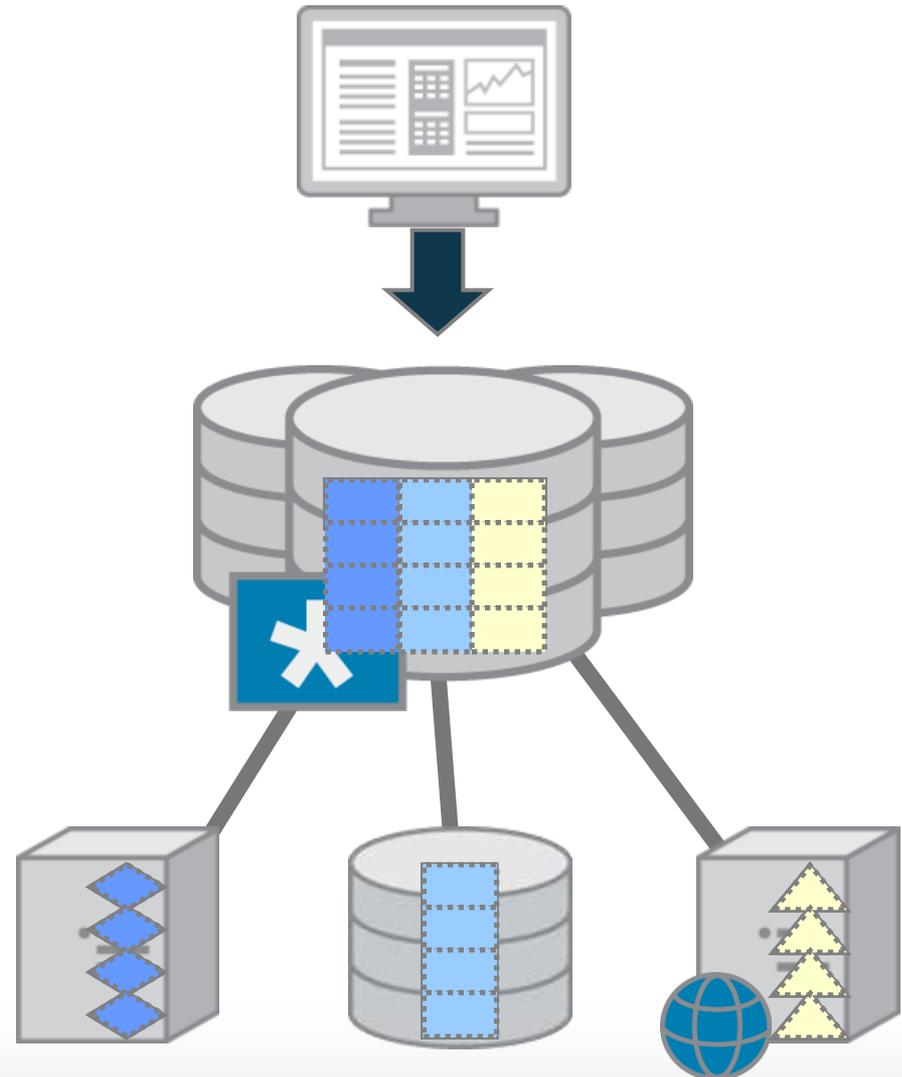
# Traditional Integration with ETL and Data Warehouses

## Traditional Approach

1. Design entire DW schema
2. Develop ETL
3. Refresh on batch basis
4. Application gets data from DW

## Issues

- Slow development cycle
- Replicated data
- Batch latencies
- Physical stores overhead



# Data Virtualization Design

## Design Steps

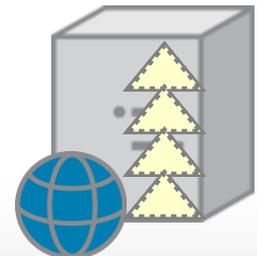
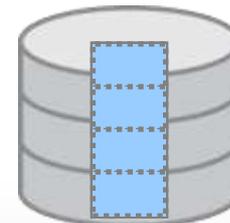
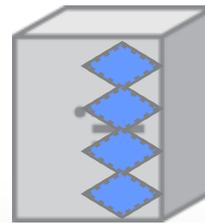
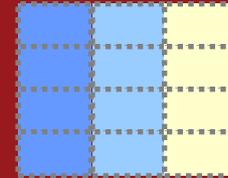
1. Discover data
2. Model individual view/service
3. Validate view/service

## Benefits

- Faster time to solution
- Easy to learn and use tools
- Extensible / reusable objects



### Data Virtualization Infrastructure



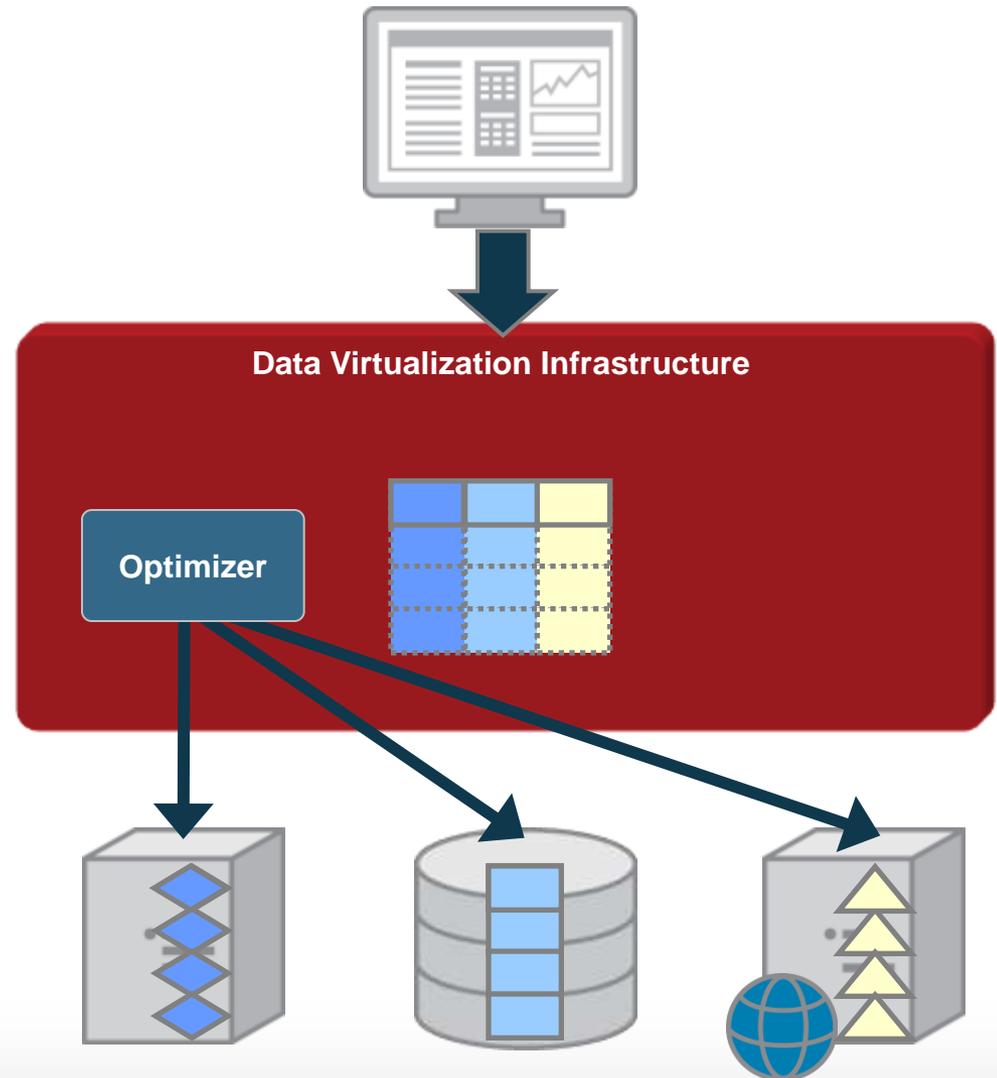
# Data Virtualization Production

## Production Steps

1. Application invokes request
2. Optimized data access and retrieval (single query)
3. Deliver data to application

## Benefits

- Less replication
- High performance
- Up-to-the-minute data



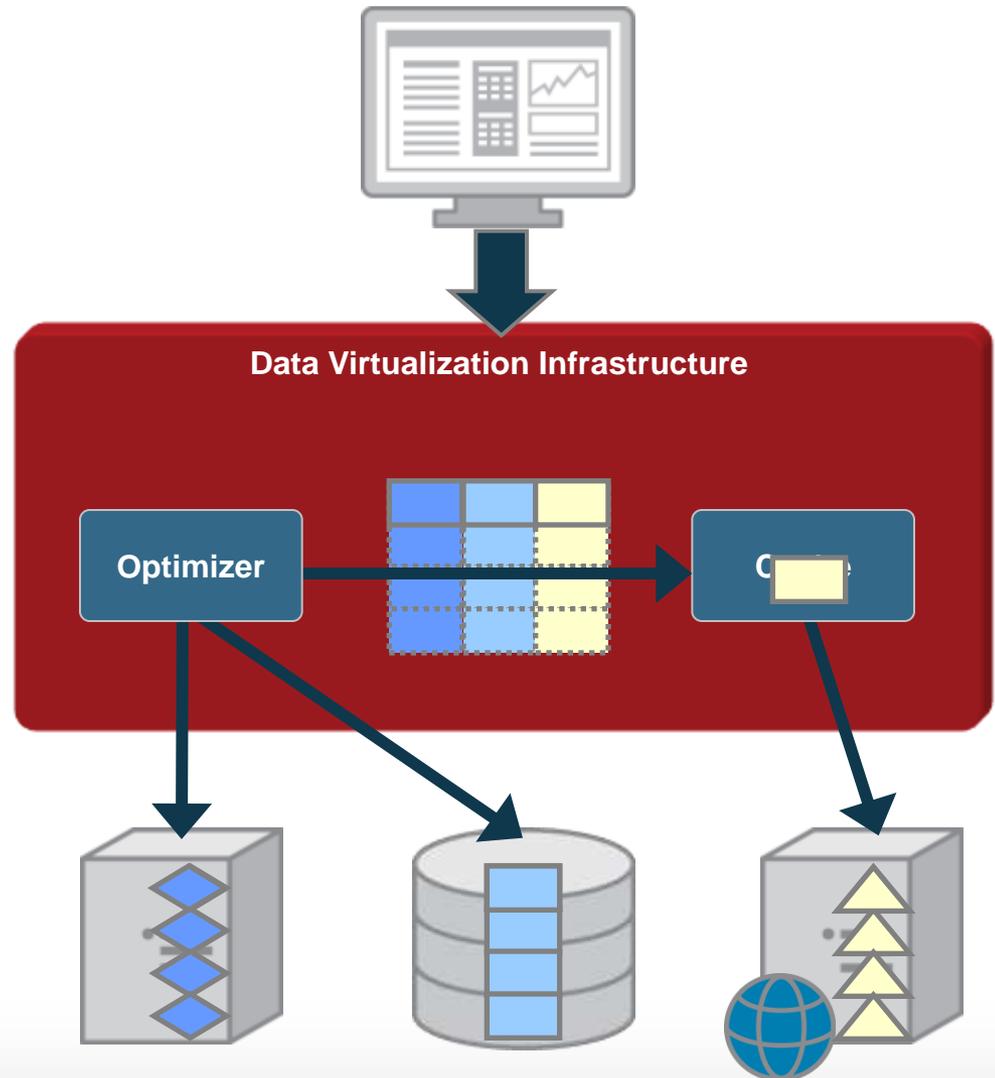
# Data Virtualization Production with Caching

## Production Steps

1. Cache essential data
2. Application invokes request
3. Optimized data access and retrieval (leveraging cached data)
4. Deliver data to application

## Benefits

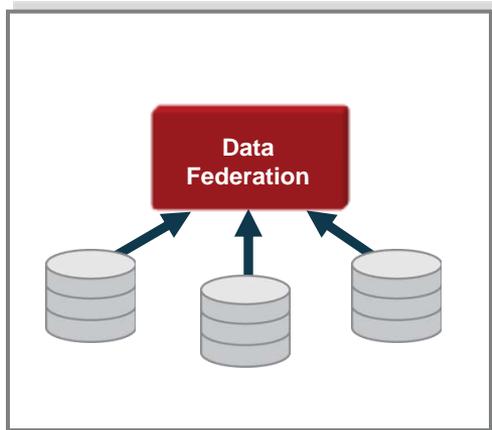
- Removes network constraints
- 7-24 availability
- Optimal performance



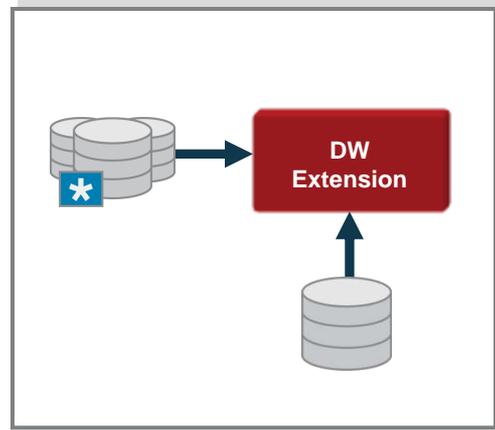
# Value Metrics: Data Virtualization vs Consolidation

Metric	To build virtual vs physical mart	Report writing
Time to Market	50-80% less time	40-90% less time
Resource Load	25% less resources	50% less resources
Object Re-use	20-30% re-use	>30% re-use

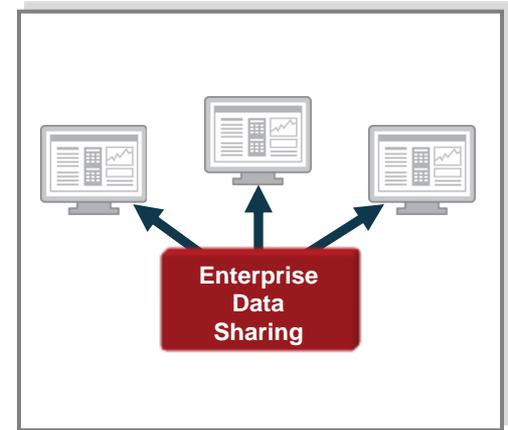
# Where Do You Use Data Virtualization?



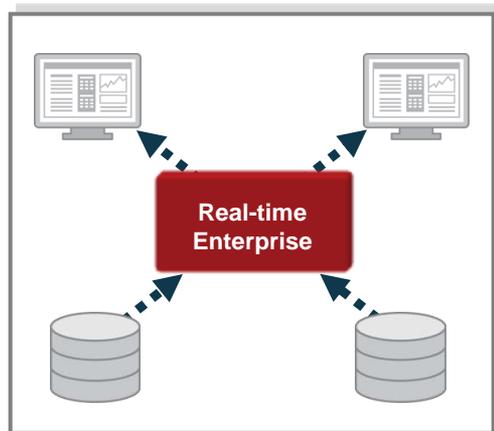
[>>Link](#)



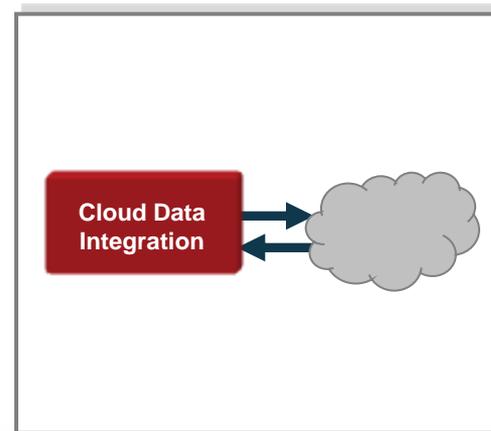
[>>Link](#)



[>>Link](#)



[>>Link](#)



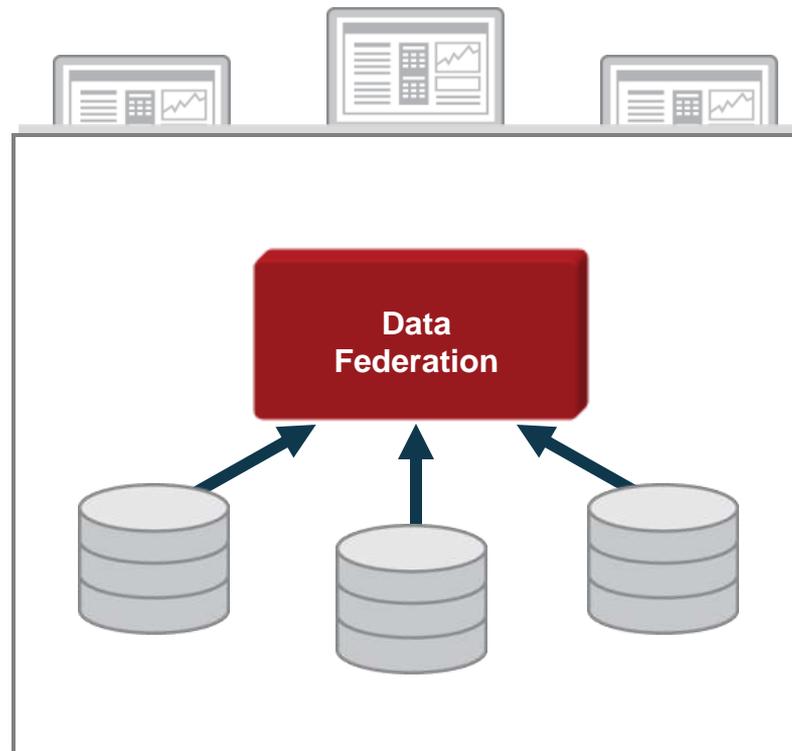
[>>Link](#)

# Data Federation

## Project Manager



“My application requires data from multiple incompatible sources.”



## Use Cases

- › [Federated Views](#)
- › [Data Services](#)
- › [Data Mashups](#)
- › [Caches](#)
- › [Virtual Data Marts](#)
- › [Virtual Operational Stores](#)

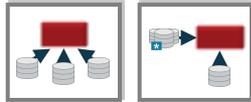
[>>Go to Technical Issues](#)  
[>>Go to Pattern Summary](#)

# Well Maintenance and Repair Energy Provider

## › Situation and Requirements:

- Needed real-time data for improving production from 10000 wells
- Complex surface, subsurface, and business data in high volumes from many disparate sources.

## › Solution:



- Build an operational data store that federates up-to-the-minute well data using Composite Information Server along with daily data warehouse data to deliver actionable information required to automate key maintenance and repair decisions throughout the day and relieve key resources for other value-adding tasks

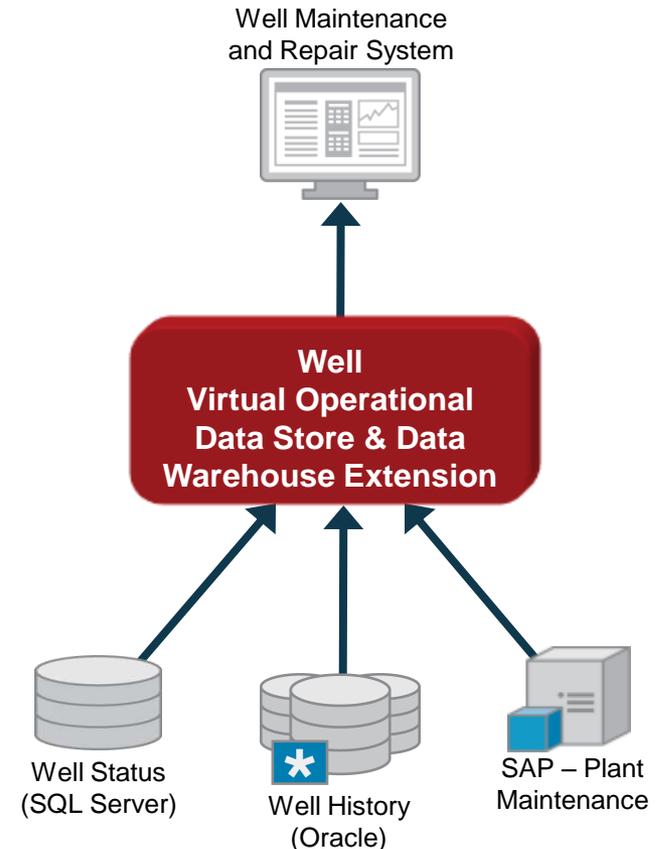
## › Results:



- 10% increase in well revenue-performance and efficiency
- 50% reduction in development time for complex integration projects
- 20% improvement in individual field engineer efficiency

*“To ‘extend the night’, we now virtualize our critical data. As a result of key process changes and an improved enterprise architecture that includes Composite Information Server, we are now operating at 98% utilization of assets. Before it was 89%”*

CIO, Large CA Energy Provider

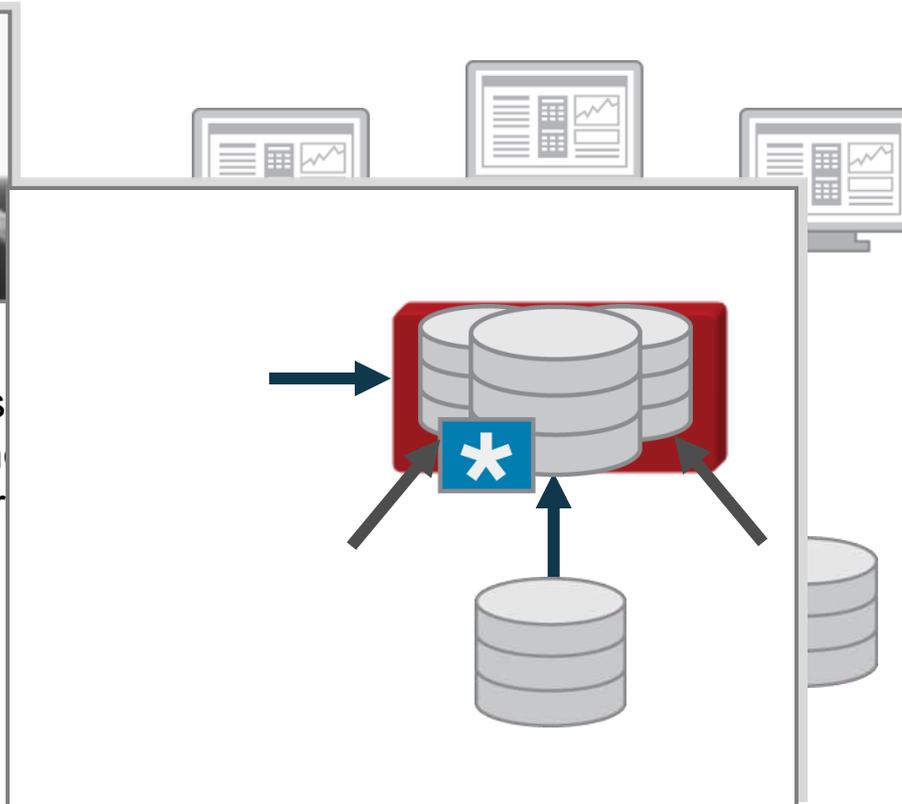


# Data Warehouse Extension

## Data Warehouse Owner



“My data warehouse does not contain all the data required for the reports we need to build.”



## Use Cases

- › [Extending Data Warehouses](#)
- › [MDM Hub Extension – 360° View](#)
- › [Data Warehouse Federation](#)
- › [Hub & Virtual Spoke](#)
- › [Enterprise Architecture](#)
- › [Complementary ETL](#)
- › [Data Warehouse Prototyping](#)
- › [Data Warehouse Migration](#)

[>>Go to Technical Issues](#)

[>>Go to Pattern Summary](#)

# Sales Management Portal Pharmaceutical

## > Situation and Requirements:

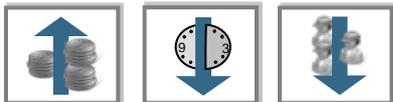
- Sales managers needed a one-stop portal to help them improve sales productivity in their regions
- Data required included highly sensitive prescriptions, sales compensation, expenses, and physician visits data

## > Solution:

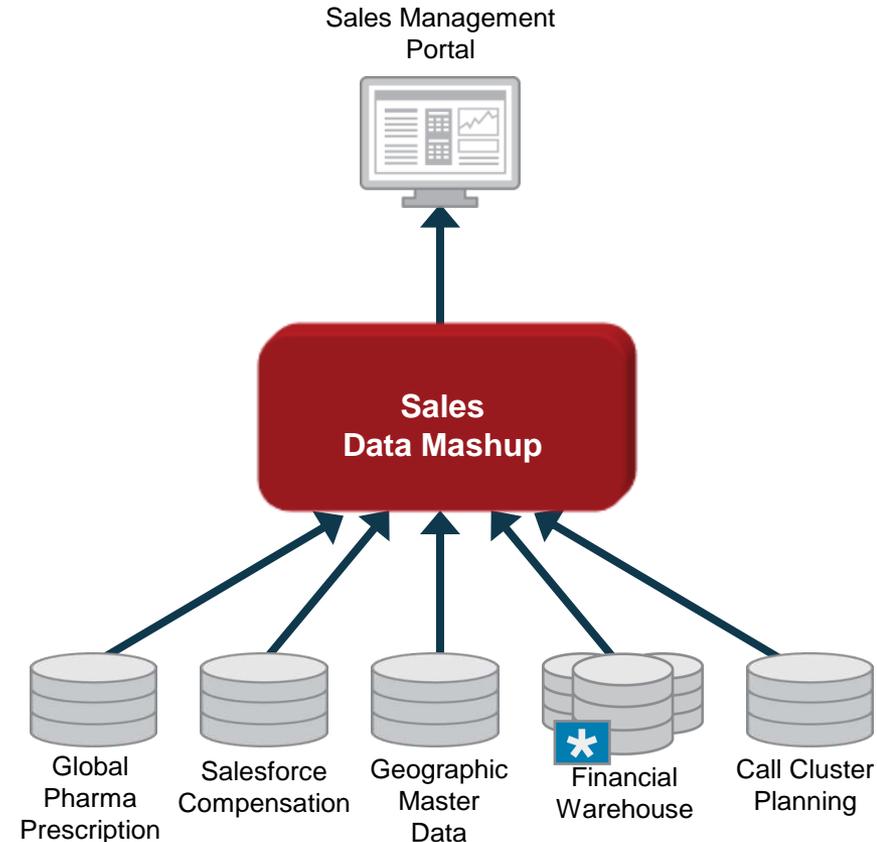


- Build sales management data mashups using Composite Information Server to federate and deliver disparate source data to sales management portal

## > Results:



- Profit increase by better sales manager visibility of sales reps performance, territory coverage, physician spend, and more
- Build new applications 25% faster, increasing flexibility to changing business and compliance requirements
- 25% cost reduction for some projects by data services reuse

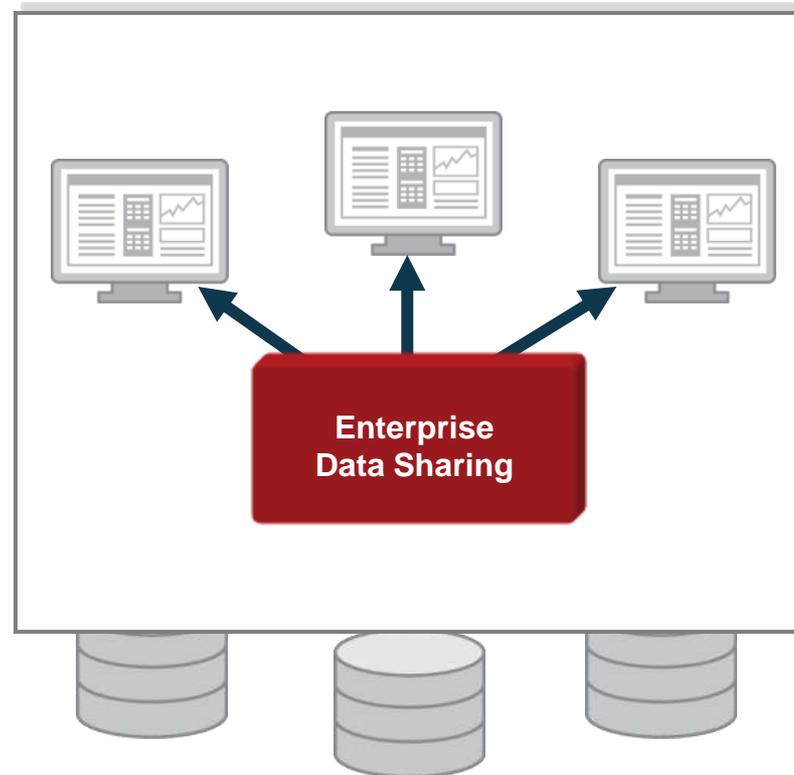


# Enterprise Data Sharing

## IT Director



“I need to share the same data across multiple applications.”



## Use Cases

- [Shared Data Services](#)
- [Data Abstraction Layer](#)
- [Standards-compliant Data Services](#)
- [Data Virtualization Competency Center](#)

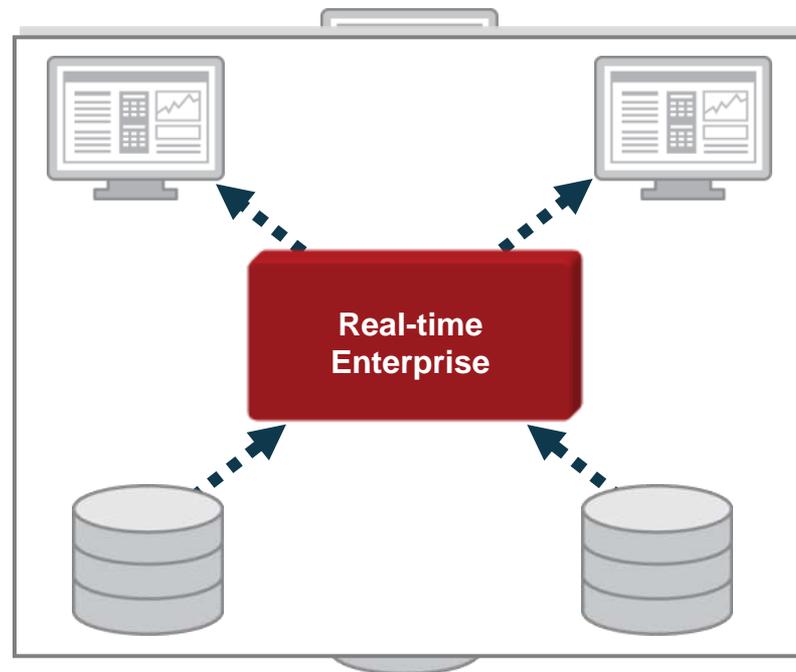
[>>Go to Technical Issues](#)  
[>>Go to Pattern Summary](#)

# Real-time Enterprise Data Infrastructure

CTO



“I need our data infrastructure to deliver the real-time information required to support our real-time business.”



Use Cases

➤ [Real-time Financial Enterprise](#)

[>>Go to Technical Issues](#)  
[>>Go to Pattern Summary](#)

# Real-time Enterprise Investment Bank

## > Situation and Requirements:

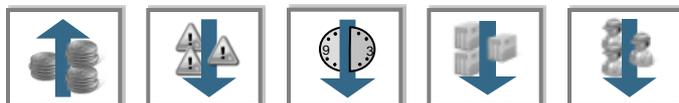
- Diminishing quality and timeliness of data
- Need for reconciliation between copies
- Need for resolution of data inconsistencies
- Complex inter-dependencies
- Lack of responsibility and authority

## > Solution:

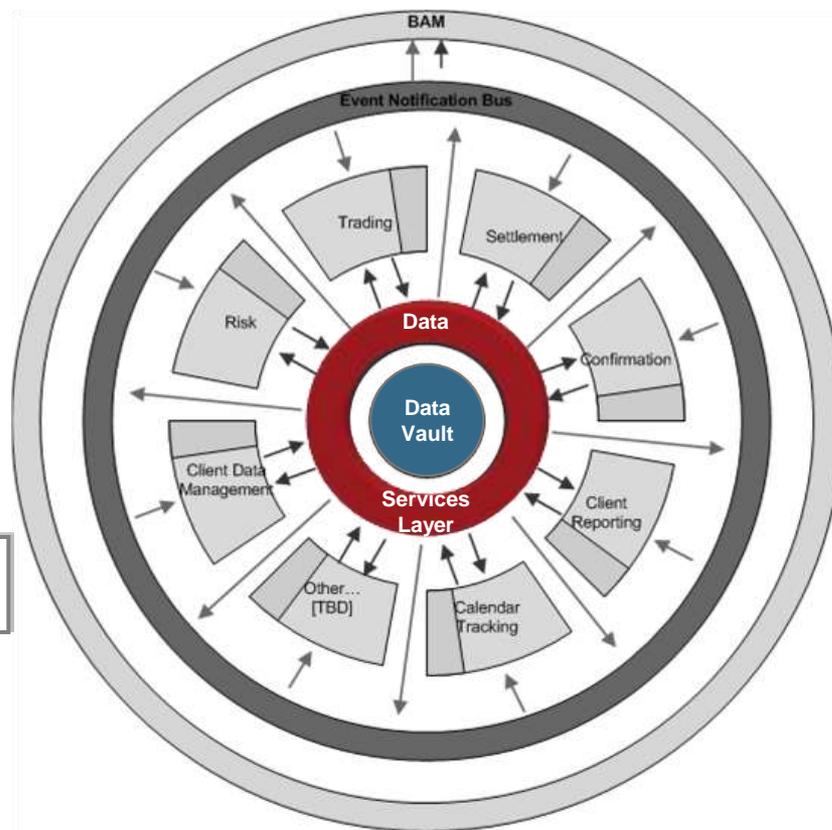


- Using Composite Information Server, built a data service layer to abstract and distribute data from a golden source of enterprise data
- Data is stored only once
- Source of data is the vault
- Each data item has only one owner
- Access is through approved data services
- Services work autonomously

## > Results:



- Improved data governance
- Increased data quality
- Reduce latency
- Less data replication
- Better, faster business operations

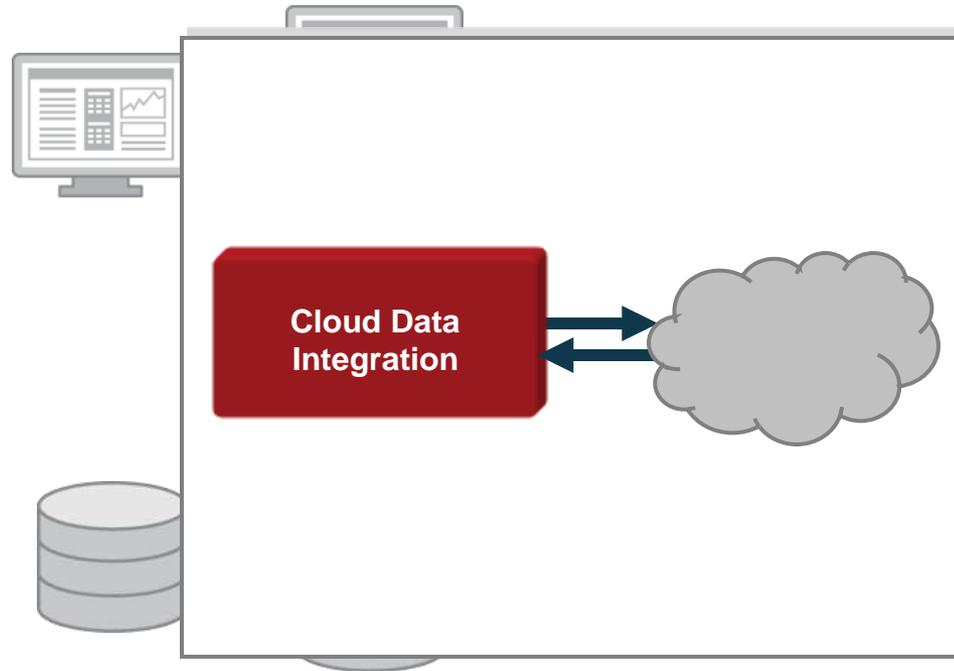


# Cloud Data Integration

CIO



“I need to integrate and synchronize data between in-house systems and applications running in the cloud.”



Use Cases

> [Salesforce.com Integration](#)

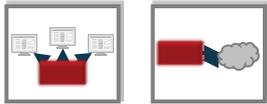
>> [Go to Technical Issues](#)  
>> [Go to Pattern Summary](#)

# Sales Management Reporting Top 5 Insurer

## > Situation and Requirements:

- Salesforce.com standard reporting not sufficient to manage the hundreds of agents
- Too costly to manually integrate Cognos reporting tools directly to Salesforce.com

## > Solution:



- Connect Cognos to Salesforce.com using Composite Salesforce.com adapter to deliver daily time series data to Cognos sales management reporting facilitated by Composite caching capabilities

## > Results:



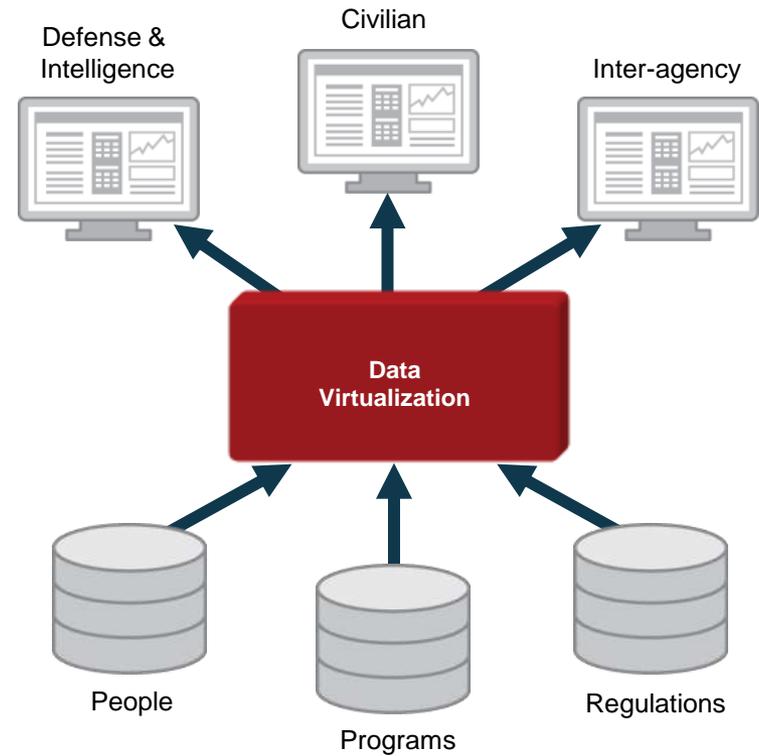
- 100% ROI in 3 months elapsed time
- 60% reduction in time to implement reports
- 80% reduction in cost to create a new report



# US Federal Government



- Inter-agency Data Sharing
- Single View of Criminal/Terrorist
- Operational Business Intelligence
- Master Data Management
- Intelligence Analyst Workbench
- Single View of Disaster Victim
- Consumer Self Service Portals



# When to Virtualize - Project Attributes

➤ “80/20” Decision Tool for Business Analysts Data Architects

➤ Project-based

- Business Considerations
- Data Source Considerations
- Data Consumer Considerations

➤ Excel and Web Versions

<http://www.compositesw.com/DIStrategyTool/DIStrategyTool.html>

Importance (Weight)			Strongly Agree	Leaning This Way	Neutral	Leaning This Way	Strongly Agree	
<b>Business Considerations</b>								
Normal ▼	<b>Time-to-Solution:</b>	My superiors and consumers are breathing down my neck to get access to this data -- I need to get this solution deployed ASAP (i.e., less than a year).	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	I can take whatever time is necessary to provide the data to my consumer consistent with my other priorities and staffing constraints.
Normal ▼	<b>Cost Sensitivity:</b>	My budget is tight, and I need to reduce infrastructure investment, programming costs, and maintenance load.	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	My budget is sufficient, and I'm not worried about infrastructure investment programming costs, or maintenance load.
Normal ▼	<b>Requirements Stability:</b>	The consumers of this data think they know what they want, but my experience tells me that as soon as I give it to them, they'll need something additional or different.	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	The consumers of this data are confident that they've thought of everything they need, and I believe their requirements are solid and stable.
Normal ▼	<b>Replication Constraints:</b>	Privacy regulations or internal policies don't allow me to replicate and consolidate data.	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	I can replicate data as often as I need to; There are no policies or regulations governing such replication.
Normal ▼	<b>Organizational Personality:</b>	My IT team prefers to apply the right tool to the right job, even when doing so means adopting and learning something new.	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	My IT team is has deep expertise in our current tool set, and they like to reduce risk by sticking to what they know.
<b>Data Source Considerations</b>								
Normal ▼	<b>Source System Availability:</b>	My operational systems are highly available, and the network is robust and reliable.	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	My source systems are often offline, and/or the network connection between them is not reliable.
Normal ▼	<b>Source System Load:</b>	My operational systems are busy, but most can handle some additional controlled access.	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	My operational systems are swamped, and I can't put any additional load on any of them during business hours.
Normal ▼	<b>Data Cleansing Needs:</b>	The data is in pretty good shape, but I need to perform some cleansing that can be done in a single pass (e.g., enrichment, normalization).	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	The data is a mess, and I need to perform data cleansing that requires multiple passes, human intervention, and should be persisted for re-use (e.g., matching, de-duping, conflict resolution).
Normal ▼	<b>Transformations:</b>	The data transformations are relatively straightforward and functional, and they can probably be accomplished by streaming data through SQL-like functionality (e.g., sub-strings, case-statements, etc.).	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	The data transformations are complex and multi-step, and I will probably need a rules engine and/or a BPM pipeline to accomplish them.
<b>Data Consumer Considerations</b>								
Normal ▼	<b>Application Focus:</b>	The consuming application provides decision-making visibility into some operational aspect of my business (e.g., current risk, inventory levels, device status).	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	The consuming application does historical analysis and/or data mining to facilitate strategic planning and long-term performance management.
Normal ▼	<b>Data Format and Access:</b>	The target data format is standard relational or hierarchical (XML); Data might be accessed via web services (SOAP).	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	The target data format is a star-schema or multi-dimensional cube; Data might be accessed via MDX queries.
Normal ▼	<b>Data Freshness:</b>	The consuming application needs current data that includes intraday changes.	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	The consuming application can work with end-of-day data from yesterday.
Normal ▼	<b>Data Volume:</b>	In order to answer each business query, I need access to a large amount of data, but individual requests read and process only a reasonably-sized set of data.	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	In order to answer each business query, I will need to read and process a large amount of data.



*Questions?*

**COMPOSITE**  
— SOFTWARE —