



Applying Data Virtualization:

13 Use Cases that Matter



DATA VIRTUALIZATION IS ON THE RISE

Analyst firms Gartner, Inc.¹ and Forrester² are projecting accelerated data virtualization adoption for both first-time and expanded deployments. What are the uses cases for this technology?

In its Data and Analytics Summits, Gartner has answered this question by identifying 13 data virtualization use cases shown here:

Traditional Analytics

- ✓ Prototyping for Physical Integration
- ✓ Data Access / Semantic Layer for Analytics
- ✓ Logical Data Warehouse Architecture
- ✓ Data Preparation

Traditional Operational

- ✓ Abstract Data Access Layer / Virtual ODS
- ✓ Registry-style Master Data Management
- ✓ Legacy System Migration
- ✓ Application Data Access

Emerging

- ✓ Cloud Data Sharing
- ✓ Edge Data Access in IoT Integration
- ✓ Data Hub Enablement
- ✓ Data and Content Integration
- ✓ Regulatory Constraints on Using Data

This paper explores each of these use cases by:

- ❖ Identifying key requirements
- ❖ Showing how you can apply **Data Virtualization** software to address these needs
- ❖ Listing the benefits you can expect when implementing **Data Virtualization** software for the use case



TRADITIONAL ANALYTIC USE CASES

I. PROTOTYPING FOR PHYSICAL INTEGRATION

Physical integration is a proven approach to analytic data integration; However, long lead times associated with physical integration — on average 7+ weeks according to TDWI — can delay realizing business value.

Further, physical integration requires significant data engineering efforts and a complex software development life-cycle. Challenges include:

- ❖ **Requirements.** Business requirements are not always clear at the start of a project and thus can be difficult for business users to clearly communicate.
- ❖ **Design.** Identifying and associating new mappings, new ETLs, and schema changes is complex. Further, current data engineering staff may not understand older schemas and ETLs. This makes detailed technical specifications a key requirement.
- ❖ **Development.** Schema changes and ETL builds are required prior to end user validation. Resultant rework cycles often delay solution delivery.
- ❖ **Deployment.** Modifying existing warehouse / data mart schemas and ETLs can be difficult and/or risky.

Prototyping physical integration using Data Virtualization software lets your data engineers:

- ❖ Interactively refine requirements, and based on actual data, build virtual data services side-by-side with business users.
- ❖ Quickly deploy agreed datasets into production to meet immediate business needs.

PHYSICAL INTEGRATION REQUIRES SIGNIFICANT DATA ENGINEERING EFFORTS AND A COMPLEX SOFTWARE DEVELOPMENT LIFE CYCLE.

- ❖ Invest additional engineering efforts on physical integration later, only if required.
- ❖ If required, use mappings and destination schema within the proven dataset as a working prototype for physical integration ETLs and schema changes.
- ❖ Once physical integration is tested, transparently migrate from virtual to physical without loss of service.

With Data Virtualization software you get:

- ❖ Faster time-to-solution than physical integration, and accelerated business benefits
- ❖ Less effort spent on upfront requirements definition and technical specification
- ❖ The right level of data engineering required to meet requirements, while avoiding unnecessary over- engineering
- ❖ Less disruption of existing physical repositories, schemas, and ETLs



2. DATA ACCESS / SEMANTIC LAYER FOR ANALYTICS

Vendor-specific analytic semantic layers provide specialized data access and semantic transformation capabilities that simplify your analytic application development.

However, these vendor-specific semantic layer solutions have limitations including:

- ❖ Delayed support of new data sources and types
- ❖ Inability to share analytic datasets with other vendor's analytic tools
- ❖ Federated query performance that is not well optimized
- ❖ Limited range of transformation capabilities and tools

VENDOR-SPECIFIC SEMANTIC LAYER SOLUTIONS HAVE LIMITATIONS

Data Virtualization software provides a vendor-agnostic solution to data access / semantic layer for analytics challenges. With Data Virtualization software you can:

- ❖ Access any data source required
- ❖ Model and transform analytic datasets quickly
- ❖ Deliver analytic data to a wide range of analytics vendor tools via industry-standard APIs including ODBC, JDBC, SOAP, REST, and more
- ❖ Share and reuse analytic datasets across multiple vendors' tools
- ❖ Automatically optimize queries
- ❖ Conform analytic data access and delivery to enterprise security and governance requirements

With Data Virtualization software you get:

- ❖ One place to go for analytic datasets regardless of analytic tool vendor
- ❖ Better analysis from broader data access and more complex transformations
- ❖ Lower costs, with reuse of analytic datasets across diverse analytic tools and users
- ❖ Faster query performance
- ❖ Greater analytic data security and governance



3. LOGICAL DATA WAREHOUSE ARCHITECTURE

Traditional data warehouses are no longer sufficient to support today's complex data and analytics landscape. The logical data warehouse (LDW) combines the strengths of traditional warehouses with alternative data management and access strategies to improve your agility, accelerate innovation, and respond more efficiently to changing business requirements.

An LDW architecture is comprehensive. It supports:

- ❖ A data services approach that separates data access from, processing from transformation, and transformation from delivery
- ❖ Diverse analytic tools and users
- ❖ Diverse data types and sources including traditional data repositories, distributed processing (big data), virtualized sources, and analytic sandboxes
- ❖ Unified business ontologies that resolve diverse IT taxonomies via common semantics
- ❖ Unified information governance including data quality, master data management, security, and more
- ❖ Service level agreement (SLA) driven operationalization

DATA VIRTUALIZATION SOFTWARE PROVIDES A VIRTUALIZATION-CENTRIC LDW ARCHITECTURE SOLUTION

Data Virtualization software provides a virtualization-centric LDW architecture solution. With Data Virtualization software you can:

- ❖ Access any source including traditional data repositories, distributed processing (big data), virtualized sources, and analytic sandboxes, both on-premises and in the cloud
- ❖ Model and transform data services quickly, in conformance with semantic standards
- ❖ Deliver data in support of a wide range of use cases via industry-standard APIs including ODBC, JDBC, SOAP, REST, and more
- ❖ Share and reuse data services across many applications
- ❖ Automatically allocate workloads to match SLA requirements
- ❖ Align data access and use with enterprise security and governance requirements
- ❖ Optionally add Master Data and Metadata software to create a more complete LDW architecture
- ❖ One logical place to go for analytic datasets regardless of source or application
- ❖ Better analysis from broader data access and more complex transformations
- ❖ Faster analysis time-to-solution via agile data service development and reuse
- ❖ Higher quality analysis via consistent, well-understood data
- ❖ Higher SLAs via loose-coupling and optimization of access, processing, and transformation
- ❖ Flexibility to add or change data sources or application consumers as required
- ❖ More complete and consistent enterprise data security and governance
- ❖ One set of master, reference and transactional data definitions, one data catalog, one point of access, one security model, and one governance system—for your entire enterprise LDW when Master Data and Metadata software are included



4. DATA PREPARATION

Self-service data preparation has proven to be a great way for business users to quickly transform raw data into more analytic friendly datasets. However, some agile data preparation needs require data engineering skill and higher-level integration capabilities.

Challenges include:

- ❖ Support for increasingly diverse and distributed data sources and types
- ❖ Limited range of transformation capabilities and tools
- ❖ Constraints on securing, governing, sharing, reusing, and productionizing prepared datasets

DATA VIRTUALIZATION
SOFTWARE PROVIDES YOU
A MUST HAVE AGILE DATA
PREPARATION SOLUTION
FOR DATA ENGINEERS
THAT COMPLEMENTS
BUSINESS USER DATA
PREPARATION TOOLS

Data Virtualization software provides you a must have agile data preparation solution for data engineers that complements business user data preparation tools. Data Virtualization software lets your data engineers:

- ❖ Interactively refine requirements and prepare datasets with business users based on actual data
- ❖ Prepare datasets that may require complex transformations or high-performance queries
- ❖ Leverage existing datasets when preparing new datasets
- ❖ Quickly deploy prepared datasets into production when appropriate
- ❖ Align data preparation activities with enterprise security and governance requirements
- ❖ Allow your citizen data engineers to easily find, modify, and author data views without calling on your technical teams via a user experience specifically designed for less-technical staff

With Data Virtualization software you get:

- ❖ Rapid, IT-grade datasets that meet analytic data needs, either as-is or as the foundation for additional data preparation by business analysts
- ❖ The right level of data engineering required to meet requirements, while avoiding unnecessary over-engineering
- ❖ Less effort spent productionizing datasets
- ❖ More complete and consistent data security and governance
- ❖ Greater value-add and smarter collaboration between your technical and citizen data engineers



TRADITIONAL OPERATIONAL USE CASES

6. ABSTRACT DATA ACCESS LAYER/ VIRTUAL ODS

Physical operational data stores (ODS) have proven a useful compromise that balances operational data access needs with operational system SLAs.

However, replicating operational data in an ODS is not without its costs. Challenges include:

- ❖ Significant development investments for ODS set up, and for integration projects that move data to them.
- ❖ Higher operating costs for managing the associated infrastructure.
- ❖ Integration workloads on the operational system.
- ❖ Often the operational source is not resource constrained, or operational queries may be light enough to not create significant workloads.
- ❖ When operational data is in an ODS, it may still require further transformations to make it useful for diverse analysis needs.

In contrast to a physical ODS, the Data Virtualization virtual ODS solution addresses these challenges. With Data Virtualization software you can:

- ❖ Access any operational data or other sources as required
- ❖ Model and transform operational datasets quickly

- ❖ Deliver data to a wide range of operational applications via industry-standard APIs, protocols, and architectures including ODBC, JDBC, SOAP, REST, and more
- ❖ Share and reuse analytic datasets across applications
- ❖ Reduce the impact on operational sources via query optimization and intelligent caching
- ❖ Conform operational data access and delivery to enterprise security and governance requirements
- ❖ Optionally add full lifecycle API management when appropriate
- ❖ One virtual place to go for operational data
- ❖ Better analysis from broader data access and more flexible transformations
- ❖ Lower costs due to less replicated data maintained in physical ODSs
- ❖ More than good enough query performance without impacting operational system SLAs





6. REGISTRY-STYLE MASTER DATA MANAGEMENT

Master Data Management (MDM) is an essential capability. Analyst firms such as Gartner have identified four MDM implementation styles (consolidation, registry, centralized, and coexistence) that you can deploy independently or combine to help enable successful MDM efforts.

Registry-style MDM implementations require:

- ❖ Access to master and reference data from diverse sources
- ❖ A cross-reference table (index) that reconciles and links related master data entities and identifiers by source
- ❖ Data services that expose the cross-reference table to analytic and operational applications that require master data from one or more sources
- ❖ Data federation that leverages the cross-reference table when querying detailed data associated with master entities

FOUR MDM IMPLEMENTATION STYLES

Data Virtualization software is ideal for registry-style MDM solutions. With Data Virtualization software you can:

- ❖ Introspect sources and identify potential master data entities and relationships
- ❖ Build a physical master data registry that relates and links master data across sources
- ❖ Cache registry copies adjacent to MDM user applications to accelerate frequent MDM queries
- ❖ Combine master, detail, and non-master data to provide more complete 360-degree views of key entities
- ❖ Optionally use Master Data software to support consolidation, centralization, and coexistence of MDM implementation styles

With Data Virtualization software you get:

- ❖ A complete solution for registry-style MDM implementations, with integrated support for consolidation, centralization, and coexistence of MDM implementation styles
- ❖ Better analysis via more complete views of master data entities across sources
- ❖ Higher analytic and data quality via consistent use of master and reference data
- ❖ Faster query performance and less disruption to master data sources
- ❖ Greater agility when adding or changing master and reference data sources



7. LEGACY SYSTEM MIGRATION

New technology provides more advanced capabilities and lower cost infrastructure. You want to take advantage. However, migrating legacy data repositories to new ones or legacy applications to new applications technology is not easy.

Challenges include:

- ❖ Business continuity requires non-stop operations before, during, and after the migration.
- ❖ Applications and data repositories are often tightly coupled making them difficult to change.
- ❖ Big bang cutovers are problematic due to so many moving parts.
- ❖ Too often, testing and tuning only happen after the fact.

Data Virtualization software provides a flexible solution for legacy system migration challenges. With Data Virtualization software you can:

- ❖ Create a loosely coupled, middle-tier of data services that mirror as-is data access, transformation, and delivery functionality
- ❖ Test and tune these data services on the sidelines without impacting current operations
- ❖ Modify the as-is data services to now support the future- state application or repository, then retest and retune
- ❖ Migrate the legacy application or repository
- ❖ Implement future-state data services to consume or deliver data to and from the new application or repository

With Data Virtualization software you get:

- ❖ To take advantage of new technology opportunities that can improve your business and cut your costs
- ❖ Loose coupling you need to divide complex migration projects into more manageable phases
- ❖ Less risk by avoiding big bang migrations
- ❖ Reusable data services that are easy to modify and extend for additional applications and users



8. APPLICATION DATA ACCESS

Your applications run on data; However, application data access can be difficult. Challenges include:

- ❖ The need to understand and access increasingly diverse and distributed data sources and types including data-in- motion and data-at-rest
- ❖ Difficulty in sharing data assets with other applications
- ❖ Federated query performance that may require optimization
- ❖ Complex transformations that may require specialized tools and techniques
- ❖ Complex data and application security requirements that need to be enforced

Data Virtualization software provides a powerful solution to these application data access challenges. With Data Virtualization software you can:

- ❖ Access nearly 350 data sources, including over 90 streaming sources
- ❖ Model and transform application datasets quickly
- ❖ Deliver data to a wide range of applications development tools via industry-standard APIs, protocols, and architectures including ODBC, JDBC, SOAP, REST, and more
- ❖ Share and reuse application datasets across multiple analytic and operational applications
- ❖ Automatically optimize queries
- ❖ Conform data access and delivery to enterprise security and governance requirements
- ❖ Optionally add full lifecycle API management when appropriate



With Data Virtualization software you get:

- ❖ One place to go for both analytic and operational application data access
- ❖ Better applications from broader data access and more complex transformations
- ❖ Lower costs from application dataset reuse across diverse applications
- ❖ Faster query performance
- ❖ Greater data security and governance



EMERGING USE CASES

9. CLOUD DATA SHARING

With the rise of cloud-based applications and infrastructure, more data than ever resides outside your enterprise. As a result, your need to share data across your cloud and enterprise sources has grown significantly. Challenges include:

- ❖ The need to understand and access increasingly diverse cloud data sources and APIs
- ❖ Diverse data consumers, each with their own data needs and application technologies
- ❖ Complex transformations that may require specialized tools and techniques
- ❖ Wide-area network (WAN) query performance that may require optimization
- ❖ Complex cloud data security requirements that need to be enforced

Data Virtualization software provides a powerful solution for these cloud data sharing challenges. With Data Virtualization software you can:

- ❖ Access nearly any major cloud data source
- ❖ Model and transform cloud datasets quickly
- ❖ Deliver cloud data to a wide range of applications development tools via industry-standard
- ❖ APIs, protocols, and architectures including ODBC, JDBC, SOAP, REST, and more
- ❖ Share and reuse cloud data across multiple applications
- ❖ Automatically optimize queries and apply caching to mitigate WAN latency
- ❖ Align data access and delivery to conform with enterprise and cloud data security and governance requirements
- ❖ Optionally add full lifecycle API management when appropriate

YOUR NEED TO SHARE DATA ACROSS YOUR CLOUD AND ENTERPRISE SOURCES HAS GROWN SIGNIFICANTLY

With Data Virtualization software you get:

- ❖ One place to go for cloud and enterprise data
- ❖ Better applications from broader cloud data access and more complex transformations
- ❖ Lower costs due to dataset reuse across diverse applications
- ❖ Faster query performance
- ❖ Greater cloud data security and governance



10. EDGE DATA ACCESS IN IoT INTEGRATION

Device data from IoT presents new analytic and operational application opportunities. Taking advantage requires:

- ❖ Directing streaming device data into edge repositories
- ❖ Understanding and accessing increasingly diverse and distributed IoT data sources and types
- ❖ Validating and enriching IoT data using non-IoT datasets
- ❖ Sharing IoT data assets across multiple analytic and operational applications
- ❖ Complex transformations that may require specialized tools and techniques
- ❖ Complex distributed edge, cloud, and data center security requirements that need to be enforced

Data Virtualization software for IoT edge data integration addresses these challenges. With Data Virtualization software you can:

- ❖ Access IoT edge data using over 90 streaming data adapters
- ❖ Transform IoT edge data using standard streaming data manipulation functions including enrichment, cleansing, and sliding windows
- ❖ Model and combine IoT data and other data sources to create integrated IoT datasets
- ❖ Deliver IoT data to a wide range of applications via industry-standard APIs, protocols, and architectures including ODBC, JDBC, SOAP, REST, and more
- ❖ Share and reuse IoT datasets across applications
- ❖ Reduce the impact on edge repositories via query optimization
- ❖ Conform IoT data access to enterprise security and governance requirements

With Data Virtualization software you get:

- ❖ One place to go for IoT edge data
- ❖ IoT datasets sooner for faster realization of IoT data opportunities
- ❖ Better IoT data, enriched with enterprise data via federation
- ❖ Lower costs via IoT data reuse across multiple applications
- ❖ Faster query performance
- ❖ Greater IoT data security and governance



II. DATA HUB ENABLEMENT

The data hub is logical architecture that enables data sharing by connecting producers of data (applications, processes, and teams) with consumers of data (other applications, processes, and teams). Master data hubs, logical data warehouses, customer data hubs, reference data stores, and more are examples of different kinds of data hubs. Data hub domains might be geographically focused, business process-focused, or application-focused.

Data hubs require that:

- ❖ The data hub provision data to and receive data from analytic and operational applications
- ❖ Hub data is governed and secure
- ❖ Data flows into and out of the hub are visible

Data Virtualization data hub solution delivers these requirements. With Data Virtualization software you can:

- ❖ Introspect sources and identify potential data hub entities and relationships
- ❖ Access any data hub data source
- ❖ Model and transform data hub datasets
- ❖ Deliver data hub datasets to diverse analytic and operational applications via industry-standard APIs, protocols, and architectures including ODBC, JDBC, SOAP, REST, and more
- ❖ Share and reuse data hub datasets across multiple applications
- ❖ Conform data hub access and delivery

to enterprise security and governance requirements

- ❖ Optionally add Master Data and Metadata software to create a more complete data hub solution

GEOGRAPHICALLY FOCUSED, BUSINESS PROCESS-FOCUSED, OR APPLICATION-FOCUSED.

With Data Virtualization software you get:

- ❖ A complete solution for data hub implementations
- ❖ Better analysis and business processes via consistent use of data hub datasets
- ❖ Higher analytic and operational application quality via consistent use of data hub datasets
- ❖ Greater agility when adding or changing data hub datasets
- ❖ Complete visibility into data hub data flows
- ❖ End-to-end data hub security and governance
- ❖ One set of master, reference and transactional data definitions, one data catalog, one point of access, one security model, and one governance system—for each data hub domain when Master Data and Metadata software are included



12. DATA AND CONTENT INTEGRATION

Content such as images, documents, recordings, and more expand your application opportunities.

Taking advantage requires:

- ❖ Understanding and accessing increasingly diverse and distributed content data sources and types
- ❖ Combining unstructured content with more traditional structured data to complete the picture
- ❖ Sharing integrated data and content assets across multiple analytic and operational applications
- ❖ Complex data security requirements that need to be enforced

The Data Virtualization solution for integrating data and content addresses these challenges. With Data Virtualization software you can:

- ❖ Model and combine data and content datasets quickly
- ❖ Deliver integrated data and content datasets to a wide range of applications via industry-standard APIs, protocols, and architectures including ODBC, JDBC, SOAP, REST, and more
- ❖ Share and reuse integrated datasets across applications.
- ❖ Conform integrated data and content to enterprise security and governance requirements

With Data Virtualization software you get:

- ❖ One place to go for integrated data and content
- ❖ Analytic and operational data enriched with image, voice, video, and other content to provide better insights
- ❖ Faster time to solution when combining data and content
- ❖ Lower costs through reuse of integrated data and content datasets across multiple applications
- ❖ Consistent data and content security and governance



13. REGULATORY CONSTRAINTS ON USING DATA

Regulatory constraints on data use continue to expand with no end in sight. These constraints include:

- ❖ Limits on what data can be seen and by whom
- ❖ The ability to anonymize data
- ❖ The ability to delete personally identifiable information
- ❖ The need to report what data you have, who has seen that data, and in what context
- ❖ Limits on moving or replicating data beyond an enterprise and/or a country

Data Virtualization software provides the controls you need to enforce regulatory constraints on data. With Data Virtualization software you can:

- ❖ Provide virtual data services that eliminate the need to replicate regulated data
- ❖ Apply data authentication, authorization, and encryption rules that conform with compliance policies
- ❖ Control access to specific rows and/or columns via granular permissions
- ❖ Implement column masking rules to hide, replace, and/or obfuscate portions of a

column's value depending on a user's level of access

- ❖ Trace source data lineage and consumer where-used
- ❖ Log all data access and user activities
- ❖ Tag and track personally identifiable information to facilitate deletion

With Data Virtualization software you get:

- ❖ Policy-driven data access and delivery that conforms to regulatory constraints
- ❖ Flexible policies and functions that help you easily adapt to new regulations
- ❖ Complete visibility into all regulatory data-related activities
- ❖ Less data replication and thus less data that requires control



CONCLUSION

As you have seen in this whitepaper, Data Virtualization software is a flexible data virtualization solution that can support myriad use cases including the 13 documented above. If your enterprise is seeking solutions for similar use cases, consider Data Virtualization software.

CONTACT US

Sydney - Melbourne - Canberra - New Zealand

Telephone: +61 2 8004 6475

Email: info@ascention.com

Website: <https://ascention.com/contact/>

1 Ziadi, Ehistham, Sharat Menon, Mark A. Beyer, and Ankush Jain. Gartner Market Guide on Data Virtualization. November 16, 2018.

2 Yuhanna, Noel. The Forrester Wave: Enterprise Data Virtualization, Q4 2017: The 13 Vendors That Matter Most And How They Stack Up. November 15, 2017.