

12 Best Practices for Data Virtualization Implementation Success

You Have a Green Light for Data Virtualization, What Do You Do Now?

You have read all the reports, evaluated the vendors, and made your decision: Data virtualization is the right investment for you.

Just one catch. Like all enterprise-grade data management middleware, before you can gain data virtualization's benefits, you first have to implement it. Successfully, that is.

For many, that will seem a significant challenge, but it doesn't have to be. If you spend the time and effort to understand best practices and lessons learned from others who have succeeded, it will help you achieve data virtualization's many benefits—and a lot sooner and with less effort than otherwise.

The purpose of this paper is to help ensure your data virtualization implementation success. Based on hard earned knowledge gained by adopters and system implementers around the world, consider this your introduction to the 12 most essential data virtualization implementation best practices.

1. Start with Your Digital Transformation Strategy.

While every journey starts with a few steps, it helps to know where you are going.

Data virtualization can support many data-driven business objectives, multiple lines of business, teams, and roles, with data that spans diverse sources and types. Given this breadth of opportunity, you must focus your efforts to achieve the biggest business impact. It is essential that you align your data virtualization implementation strategy with your digital transformation strategy.

Will your data-driven success be realized by better customer engagement, faster innovation, operational excellence, or some combination? The answer to this question informs your data virtualization implementation strategy.

Will your data-driven competitive success be realized by better customer engagement, faster innovation, operational excellence, or some combination? The answer to this question informs your data virtualization implementation strategy.

For example, if your digital transformation strategy is to increase cross-sell revenues within your installed base, then focus your data virtualization efforts on building 360 degree views of your customer data. Knowing your customers' preferences, prior purchases, current activity, and more provides the insight you need to intelligently offer just the right cross-sell product at just the right time, and do so in a way that engages and delight them.

Over time, your digital transformation strategy and data-driven response will evolve and grow, and, so, too, will your data virtualization implementation. The good news along the way is that data virtualization's ability to do many things well and quickly will accelerate your transformations.

2. Create a Comprehensive Implementation Plan.

Successful data virtualization implementations can be complex. It helps to have a comprehensive plan that covers your journey from project kickoff to steady-state operations. Be sure your implementation plan includes these elements:

- Strategy and Architecture
- Installation Setup and Configuration
- System Administration
- Design and Development
- Operations Management and SLAs
- Continuous Improvement

To improve your odds of success, engage system integrators (SIs) with strong data virtualization practices and positive references. These specialists have many assets that can help you in several significant ways.

- First, you can take advantage of their knowledge gathered from across many implementations. These let you avoid common pitfalls and the wasted effort of reinventing the wheel. Project work breakdown structures are one example of the kinds of proven best practices you can reuse.
- Second, SIs can complement your people with theirs, filling your short-term need for extra staff required to perform the extra work associated with the initial implementation.
- Third, because they are data virtualization domain experts, they can serve as interim subject matter experts until the time when your team is fully skilled up.

3. Align with Your Data Architecture Strategy.

Rationalizing today's increasingly distributed data landscape is a big reason why you decided you needed data virtualization in the first place. Logical data warehouse, data fabric, and data-as-a-service are typical labels for these new types of modern, distributed data architectures, and data virtualization is a common element in all of them.

Data virtualization's consistent, secure, and governed common data layer makes it easy for your business users to find and use all your most important data regardless of location. And it provides the decoupling opportunity that lets your IT team store and manage your data, independently from how it is consumed. So it stands to reason that your data virtualization implementation strategy aligns to your data architecture strategy.

But just as Rome was not built in a day, neither is a distributed data architecture. As you transition from your current to your future state of data architecture, you will need to execute some difficult migrations. Using data virtualization as a migration tool lets you insulate your consumers as you move their data sources, insulate data sources as you migrate your consumers, or do both depending on how your data architecture evolves. For more detailed tips and tricks on how you can use data virtualization to accelerate these architectural driven migrations, [read this blog](#).

4. Start with Best Fit Use Cases.

Data virtualization is not the best solution to every data integration challenge. Sometimes it is, but sometimes it's better to pre-consolidate your data in a warehouse or mart—and sometimes a combination of integration methods is your best choice. Getting this decision right for each data integration project is not only important to the success of that project, those successes also influence your organization's perception of data virtualization overall. And a positive reputation based on a positive start will help you garner the executive support and additional resources you need for long-run success. So it is important to select your use cases wisely.

How do you know which use cases are the best fit for data virtualization? There are several best practices you can use to answer this question.

Examine use case attributes

Because many factors can influence whether or not to use data virtualization, you need to intelligently consider all these factors in a consistent and systematic way. TIBCO provides a [Data Integration Strategy Recommendation Tool](#) that lets your architects, designers, developers, and/or project leaders make more informed decisions on when to virtualize, when to consolidate physically, and when to apply a hybrid combination.

Synthesized from conversations with industry analysts and hands-on project teams—and now adopted by TDWI for their data virtualization courses—this tool applies 13 decision factors as follows:

- **Business Considerations**
 1. Time-to-Solution
 2. Cost Sensitivity
 3. Requirements Stability
 4. Replication Constraints
 5. Organizational Personality
- **Data Source Considerations**
 6. Source System Availability
 7. Source System Load
 8. Data Cleansing Needs
 9. Transformations
- **Data Consumer Considerations**
 10. Application Focus
 11. Data Format
 12. Data Freshness
 13. Data Volume

Pattern-match common data virtualization use cases

In some cases the best way to decide on whether to use data virtualization is to compare and contrast your current use case against proven data virtualization use cases. When you identify a match, you can proceed confidently. [Applying Data Virtualization: 13 Use Cases that Matter](#) is a useful resource to help you compare and contrast your use cases with the most frequently adopted data virtualization use cases. For each case, the paper:

- Describes what the use case does
- Identifies key business and technical challenges that must be addressed
- Explains how you can apply data virtualization to support these needs
- Lists the data virtualization benefits you can expect

Virtualize first

If you are like other data virtualization adopters, the longer you use data virtualization, the less you will rely on the strategy recommendation tool and pattern matching approaches above. You will simplify your decision process by applying data virtualization as your default method of data integration for every new use case request. And you do so even when you know that physical data consolidation may be the ultimate implementation method.

Why virtualize first? Your benefits from a virtualize-first approach include:

- **Faster response to business needs** – Virtualized datasets are far quicker to build and iterate than physically consolidated datasets, so your business gets the new data it needs far sooner.
- **Higher data engineering productivity** – Virtual integration requires less data engineering than physical integration. Built-in intelligence also reduces data engineering development and performance tuning workloads. This saves time and money. Lots of it.
- **Smarter resource use** – Use of less agile, higher cost data consolidation with ETL becomes the exception rather than the rule.
- **Better ETL** – And in the subset of cases when ETL is the appropriate final solution, the data virtualization working prototype you used to support the business initially greatly simplifies your ETL development effort and improves its quality.

5. Adopt in Stages.

As noted above, your data virtualization implementation can support many diverse use cases from simple projects to complex enterprise-wide data architectures. When faced with so many opportunities, what is the best way for you to get started and then expand to broader use?

Based on knowledge gained from hundreds of data virtualization implementations, staged deployment and adoption has proven the best practice. The key to success is start small, build your skills, and then grow based on proven results.

Stage 1: Use Case Scale

Implementing a series of small projects is your best starting point. This lets you provide the virtualized datasets you need to quickly deliver business benefits, while avoiding the risk inherent in large projects. Because you are new to data virtualization, keep your implementations simple. Do not over engineer the views and services you build. At this stage be “good enough for now,” knowing you can improve these assets in the next stage when you have expanded your team, refined your skills, and developed your best practices.

Stage 2: Domain Scale

With the winds of success at your back, you can expand your data virtualization scope. Instead of focusing only on the datasets you need for presenting use cases, you can think more broadly. This is the time to develop your standards and engineering best practices, perhaps in the context of a Center of Excellence. And this is when you gather reusable virtualized datasets you can use to support larger domains, for example, the dataset you need for:

- A specific line of business, such as North American operations
- A function, such as sales or R&D
- A class of data, such as customer related data

Stage 3: Enterprise Scale

Eventually you can pull together multiple domain-scale implementations into a broader enterprise-wide data virtualization implementation. This is the stage where you:

- Make the engineering investments required to support massive scale
- Realize your future stage data architecture
- Implement your enterprise data model

But recognize that your definition of enterprise scale can be different than other organizations. For example, enterprise data virtualization at one of the largest banks in the world spans seven business domains including high net worth customers, credit card customers, back office operations such as security services, fixed income risk management, and more. Yet significant lines of business are still in earlier stages.

6. Leverage a Center of Excellence Approach.

Data virtualization centers of excellence (CoE) are a proven people, process, and technology approach you can use to build competency, drive consistency, and accelerate value realization. When implementing a data virtualization CoE, you have two options you can choose based on your organization's unique situation and needs.

- Option 1 is to **extend an existing CoE**, for example your Data Integration Competency Center of Excellence. This option lets you leverage your existing CoE infrastructure, organizational alignment, best practices, and lessons learned.
- Option 2 is to **create a specialized, standalone CoE** that exclusively supports your data virtualization implementation.

Either way, your goal remains the same with multiple benefits and results including:

- Guide your overall data virtualization strategy and architecture to strengthen business alignment and drive greater adoption.
- Concentrate subject matter expertise to resolve data virtualization design and development issues faster.
- Reuse data virtualization objects to improve data quality and consistency, as well as save development time and resources.
- Optimize internal and external subject matter experts across multiple data virtualization projects and activities.
- Optimize shared data virtualization infrastructure utilization.
- Develop and share best practices that optimize your work.
- Better anticipate and proactively overcome obstacles to avoid extra costs.

So you can fully focus your early CoE efforts, start small by limiting the scope of data virtualization CoE functions, projects, or scope in some way. As you achieve results, communicate them to reinforce the CoE team's successes. Then use the momentum you gain to accelerate further implementation success.

7. Progress toward Your Enterprise Data Model.

One of data virtualization's greatest business benefits is its ability to shield your data consumers from your data complexity. Delivering business-friendly data that is easier to understand and use, lets you gain the data-driven insights you need to compete and win.

But you cannot achieve this with data virtualization alone. You need to combine your data virtualization implementation efforts with your enterprise data model management efforts.

The Adopt in Stages and Leverage a CoE Approach best practices above describe key process elements you can use to align the timing and scope of your common data virtualization and data modeling efforts.

The data virtualization abstraction best practices summarized below describe the data layer architecture you can use to instantiate your enterprise data model. Its three primary layers include:

- **Application Layer** – In the Application Layer, you model virtualized views and services that meet your end consumers' needs. Building on views and services in your Business Layer, this is where you refine your data for specific consumers. This might require:

- Filtering the data to select only what your consumers are entitled to see
- Reformatting the data from relational to a Web service, so it works better with your consumers' applications
- Localizing attribute names to better match terminology your consumers understand
- **Business Layer** - The Business Layer is where you instantiate your enterprise data model, leveraging and transforming views and services from the Physical Layer. It assumes you have a standard or canonical way of describing key business entities, such as customers and products, and high usage entities such as reference data. Typically, your data modelers work with your business experts and data providers to define this common set of "logical" or "canonical" views that represent your most important and commonly used business entities.
- **Physical Layer** - The Physical Layer is where your physical data sources are modeled using data source metadata. Value added tasks such as name aliasing, value formatting, data type casting, derived columns, and light data quality checks, are also performed in this layer.

Across your data virtualization enterprise data model implementation, governance is critical. Ensure your data model governance roles and responsibilities extend into your data virtualization CoE. Mandate a data model stewardship process that maintains consistency using shared workflows. Implement a common change process that ensures implementation success.

For more detail on these modeling best practices, see this [whitepaper](#).

8. Design for Performance.

When you implement data virtualization for the first time, you may be concerned about query performance, possible production system impacts, and how it aligns with your existing data integration and storage strategies.

Questions that typically arise include:

- Will using data virtualization meet your query performance objectives?
- How will it impact your production systems?
- What if someone tries a "SELECT * everything" query?
- Can you continue to use your data warehouses, marts, and lakes, or do you need to virtualize everything?
- Aren't you better off continuing to use ETL and physical consolidation, just in case?

This whitepaper, [10 Best Practices for Using TIBCO Data Virtualization to Achieve Performance SLAs](#), summarizes the essential design for performance and workload management best practices you can use to design for optimized data virtualization performance. Here is the list:

- 1 Keep your eyes on the prize.
- 2 Continue with what already works.
- 3 Understand when best to virtualize versus physically consolidate.
- 4 Just do it and see what happens.
- 5 Let built-in intelligence optimize performance for you.
- 6 Leverage massive parallel processing for an extra performance boost.
- 7 Manually tune your queries for corner cases.
- 8 Apply caching only when it makes sense.
- 9 Implement workload management as an SLA insurance policy.
- 10 When all else fails, fall back on alternative data integration techniques.

9. Consider Security from the Start.

One of the biggest benefits of data virtualization is how it enables you to consistently implement data security across your data sources and consumers. With it, you can ensure that your authenticated users can see only appropriate, authorized data and nothing more. And when even higher levels of security are needed, data virtualization can mask and encrypt your data. This [whitepaper](#) describes these capabilities in more detail.

Because security impacts every view and data service you build, every data source you query, and every user who consumes these datasets, you must consider data security holistically across every stage of your data virtualization implementation. For best success, apply these best practices:

- During your initial data virtualization server set up, establish a small group of security administrators to manage your data virtualization security in a holistic way.
- Your administrators can then set up your developers and data consumers to control who can perform what development activities as well as who can access what data. Intelligent use of security policies can help your administrators implement data virtualization security faster and more consistently.
- During development, data virtualization security automatically enforces development resource ownership, sharing, and modification. Test that these are being applied as you expect. Modify if necessary.

- During execution, data virtualization security automatically enforces user authentication and access authorization. Again, test that these are being applied as you expect. Modify if necessary.
- When you implement new data virtualization releases, evaluate how you can take advantage of the additional security capabilities including new security standards support.

10. Don't Forget Data Quality.

Your data virtualization implementation will force you to drill down into your data. As a result, you will uncover many data quality challenges. But the good news is you can use data virtualization several ways to directly improve your data quality.

Data virtualization's centralized metadata-driven syntactic and semantic transformations and enrichments let you standardize your datasets and encourage reuse. This means everyone is on the same page, improving not only the quality of data itself, but every application where it is used.

Data virtualization also affords you the opportunity to make fewer copies of your data. Fewer copies improves your data quality by substantially reducing the transformation errors and entropy inherent in typical multiple-copy data warehouse and data lake deployments. For more on how data virtualization lets you create fewer copies, read [To Improve Data Quality, Stop Playing the Data Telephone Game](#).

And when things change, as they inevitably do, it's a lot easier for you to modify data virtualization's centrally managed metadata definitions than it is to modify the multiple distributed ETLs and database schemas you need when using physical data integration. So your data quality doesn't degrade with every business change.

11. Unify Your Data Management Tools.

If yours is like most organizations, your data management tools include a broadly diverse set of sometimes adjacent, sometimes overlapping, vendor applications that span metadata management, master data management (MDM), data governance, data catalog, data modeling, data quality, and data security — as well as the various data integration types including replication, bulk/batch ETL, and data virtualization. Often purchased at different times for distinct purposes, these siloed data management capabilities increase complexity, costs, and organizational friction.

While data virtualization provides many of the core metadata management, modeling, quality, and governance capabilities you require, it does not cover everything you need. But it can be a great foundation for better synergies with other data management solutions. Doing this right will let you achieve better data management success with less effort.

The first step is to embrace today's data management functionality convergence. When selecting your data virtualization vendor, look for those who also provide advanced data management capabilities in converged platforms you can use to unify your data management capabilities. These include metadata management, MDM, data governance, data quality, data catalog, data modeling, data security, and other data integration methods. This will simplify your data management silo problem, while saving time, cutting costs, and reducing friction.

The second step is to simplify your approach to data management. Have you ever wondered why you have treated your metadata, master data, reference data, transaction data, and streaming data assets as unique, unrelated data types and manage them in unique, unrelated ways? You no longer have to. Data virtualization lets you unify what is common across your data assets including how they are accessed, modeled, secured, governed, and more. This holistic approach makes it easier to manage your data assets, especially your most important, frequently used ones.

12. Press Onward.

Once you complete your initial implementation, what happens next? Do you rest on your laurels or push ahead pressing your data virtualization fueled advantage? With so many new use cases and valuable benefits now accessible, and new data virtualization capabilities coming your way with each new release, pressing onward is the best answer.

Pressing onward requires stamina. A stamina fueled by achievement as success begets success. But first you must communicate your successes.

- Take the time to showcase your use cases and their business benefits. Others with similar use cases will knock on your door.
- Celebrate the great work performed by your implementation team members. Patting them on the back helps these internal evangelists work even harder.
- Host lunch-and-learns at your data virtualization CoEs. This makes it easy for those new to data virtualization to get on board.

And keep pace with the ever changing world of data virtualization, so you can take advantage of new opportunities as they arise. To stay informed:

- Join your data virtualization vendor's user community, get on their mailing lists, and follow their announcements so you can mine their knowledge bases.
- Follow data virtualization thought leaders such as [Rick van der Lans](#) on Twitter and LinkedIn. And while you are at it, monitor [#data_virtualization](#) and [#datavirtualization](#) on both. This broad exposure helps you see greater possibilities.
- If you subscribe to analyst firms such as Gartner, Forrester, and IDC, you can read their latest data virtualization research reports as well. These forward looking syntheses help you see where data virtualization is going in the next three to five years.

Conclusion

Implementing data virtualization can be a significant challenge. But one made easier if you master the best practices and lessons learned from others who have succeeded. This paper provides 12 implementation best practices you can apply to help ensure your data virtualization implementation success.

Take advantage, so you can make data virtualization your organization's cutting-edge advantage.

Cheers to your success.



Global Headquarters
3307 Hillview Avenue
Palo Alto, CA 94304
+1 650-846-1000 TEL
+1 800-420-8450
+1 650-846-1005 FAX
www.tibco.com

TIBCO Software Inc. unlocks the potential of real-time data for making faster, smarter decisions. Our Connected Intelligence platform seamlessly connects any application or data source; intelligently unifies data for greater access, trust, and control; and confidently predicts outcomes in real time and at scale. Learn how solutions to our customers' most critical business challenges are made possible by TIBCO at www.tibco.com.

©2021, TIBCO Software Inc. All rights reserved. TIBCO and the TIBCO logo are trademarks or registered trademarks of TIBCO Software Inc. or its subsidiaries in the United States and/or other countries. All other product and company names and marks in this document are the property of their respective owners and mentioned for identification purposes only.

27Apr2021