Composite Data Virtualization

Data Virtualization Implementation Methodology and Best Practices

Composite Software

April 2011
# TABLE OF CONTENTS

**INTRODUCTION** .................................................................................................................................... 3

**STRATEGY, PLANNING AND GOVERNANCE** ................................................................................... 4
- Establish Data Virtualization Strategy and Policies .......................................................... 4
- Specify Project Portfolio .................................................................................................. 4
- Identify Roles and Responsibilities ..................................................................................... 4
- Plan Staff Development and Education ........................................................................... 5
- Establish Data Virtualization Governance and Control ...................................................... 5

**SOLUTION ARCHITECTURE** ................................................................................................................ 6
- Define Future State Architecture ...................................................................................... 6
- Evaluate Current State Architecture ................................................................................. 6
- Develop a Plan to Migrate from the Current to a Future State .............................................. 7
- Align Specific Projects within Overall Architecture ....................................................... 7
- Identify and Resolve Issues ................................................................................................. 7

**CONFIGURATION & SET UP** ................................................................................................................ 9
- Define Data Virtualization Environments ........................................................................... 9
- Install Composite Data Virtualization Products ............................................................... 9
- Set up Security .................................................................................................................. 9
- Set up Access to Data Sources and Consumers .............................................................. 10
- Integrate Development and System Management Tools and Methods .......................... 10

**DESIGN & DEVELOPMENT** ................................................................................................................ 11
- Design Required Views, Services, Caches, etc. ................................................................. 11
- Develop ............................................................................................................................... 11
- Test and Validate ............................................................................................................... 12
- Tune Performance ............................................................................................................. 12
- Document Views and Services .......................................................................................... 12

**DEPLOYMENT** ..................................................................................................................................... 13
- Identify Service Levels and Management Requirements .................................................. 13
- Develop High Availability and Disaster Recovery Policies and Processes ......................... 13
- Deploy Completed Objects on Downstream Environments ............................................. 13
- Document Deployment Processes and Deployments ....................................................... 13

**OPERATION** ........................................................................................................................................ 14
- Ensure Reliable Operation ................................................................................................. 14
- Administer Data Virtualization Environments and Users .................................................. 14
- Tune Performance ............................................................................................................. 14
- Resolve Problems ............................................................................................................. 15
- Implement Patches ............................................................................................................ 15

**IMPROVEMENT** ................................................................................................................................... 16
- Implement New Composite Releases .................................................................................. 16
- Adopt Additional Composite Features ............................................................................... 16
- Consider New Use Cases .................................................................................................... 16
- Develop Reuse Strategies .................................................................................................. 17
- Improve Run-Time Operations .......................................................................................... 17

**CONCLUSION** ...................................................................................................................................... 18
INTRODUCTION

Composite’s proven Data Virtualization Implementation Methodology, which incorporates Best Practices, is compiled from our successful experiences across hundreds of customer deployments and ensures the following:

- Your data virtualization objectives are achieved efficiently and effectively;
- Your data virtualization projects have the best chance of success; and
- You gain the highest possible return on your investment in data virtualization.

The Composite Data Virtualization Implementation Methodology and Best Practices are designed to integrate with your existing design, development and deployment processes. When blended with internal and external staff resources, this combination of knowledge, tools, processes and more enables both immediate implementation success and on-going self-sufficiency.

The Methodology spans the complete software development lifecycle. Major stages of the Methodology include:

- Strategy, Planning & Governance
- Solution Architecture
- Configuration & Set Up
- Design & Development
- Deployment
- Operation
- Improvement

This paper summarizes details key activities and resources to successfully navigate each stage including documentation, white papers, tools, consulting services and more. Armed with the proper knowledge and support, you will be able to avoid pitfalls and efficiently and effectively work through the implementation tasks.

Please note that many of the documents listed are available only to customers through Composite Support. If you are a prospective customer and wish to examine any of these materials, please ask your sales representative for assistance.
In the Strategy, Planning and Governance stage, you develop the framework for accomplishing your overall data virtualization objectives.

**Establish data virtualization strategy and policies**

Establish your data virtualization strategy and usage policies by first understanding what data virtualization has to offer. Learn how data virtualization is used at other organizations including general usage patterns and specific project use cases.

- **Composite Data Virtualization Usage Pattern White Paper** – This white paper introduces the most popular data virtualization usage patterns, the business and IT challenges behind them and the solution that Composite delivers to overcome these challenges. Use it to help set your data virtualization adoption strategy.

- **Composite Customer Case Studies** – These case studies detail Composite data virtualization use cases. Use them to further identify specific data virtualization opportunities.

**Specify project portfolio**

Identifying specific data virtualization projects is a matter of technical fit and priority. Composite provides tools to help with both.

- **Composite Platform Technical Overview** - This white paper offers a technical overview of the Composite’s Data Virtualization Platform. Use it to understand technical capabilities you can bring to bear on your data virtualization projects.

- **Data Integration Strategy Decision Tool** – The Composite Data Integration Strategy Recommendation Tool enables you to quickly evaluate data integration options (Data Virtualization, Physical Data Consolidation, or Hybrid Combination) on a project-by-project basis. Use it to guide and reinforce your data integration middleware decisions.

- **Data Virtualization Opportunity Assessment Tool** – This tool provides the structure needed to assess multiple data virtualization opportunities relative to one another. Use it to help you organize and prioritize your entire data virtualization project pipeline including project owners, level of difficulty, and potential return on investment.

**Identify roles and responsibilities**

Along with strategy, it is also important to develop the right team. Each project requires subject matter expertise, technical skills, and the proper blend of internal personnel versus external consultants. Detailing the job description and properly separating roles and responsibilities also helps set the tone for high morale and productivity.
Plan staff development and education
Create a multi-faceted plan to educate your wide range of IT staff on data virtualization. Early on, higher level training will enable your team to see the capabilities of data virtualization. As you take on projects, more detailed training will be required. Throughout the implementation process, hands-on development with expert consultants from Composite and partner system integrators will help your staff internalize data virtualization skills. This up-front investment facilitates the knowledge transfer process, saving time and promoting efficiency in the long run.

- **Composite Training Catalog** – This catalog lists all training programs available through Composite Software from one-day primers to four-day courses. Use this to develop staff development plans.

Establish data virtualization governance and control
Data governance policies and controls should prevent undesired activities from occurring as well as trace possible improper activities after the fact. Proper security configuration (discussed later) in alignment with your authentication, authorization and encryption policies is critical. Further, auditing requirements including data lineage and transaction logging must be designed and deployed.

- **Data Virtualization Governance Advisory Service** – Composite Professional Services has a unique understanding of the intersection of governance and data virtualization. Use this service to help you establish appropriate data virtualization governance policies and controls in alignment with your broader IT and data governance policies and controls.
In the Solution Architecture stage, you identify your future-state data virtualization architecture and establish your plan to get there.

**Define future state architecture**

Your future state data virtualization architecture must consider many factors including:

- Today’s business requirements and agility to meet new requirements that will certainly arise;
- Flexibility to accommodate both existing applications as well as new applications that may come online down the road; and
- Incorporation of open standards within your architecture to “future proof” your environment as much as possible.

Composite provides several resources to help you envision a future state data virtualization architecture that fits your needs.

- **Data Virtualization Architecture Advisory Service** – Composite Professional Services has a number of architects with unique data virtualization architecture design skills. Use this service to help design your future state data virtualization architecture to meet your long term objectives.

- **Eight Ways Composite Data Virtualization Adds Value to Enterprise Data Warehousing White Paper** – This paper describes eight specific integration patterns that combine both enterprise data warehouses and data virtualization. Each pattern includes the data warehousing challenge, the enterprise data warehouse and Composite data virtualization combined solution, and example use cases. Use this to integrate data warehouses into your data virtualization architecture.

- **Composite Data Abstraction Best Practices White Paper** – This paper outlines how Composite data virtualization is an optimal way to implement an abstraction layer at enterprise scale. From an enterprise architecture point of view, Composite’s data virtualization solution provides semantic abstraction via a data services layer in support of multiple consuming applications. Use these guidelines to understand the canonical modeling objectives and levels of abstraction within your future data virtualization architecture.

**Evaluate current state architecture**

The current state of your information architecture can be complex. Evaluate it relative to the following considerations:
• Current inventory of data sources, integrations, and consumers;
• Current reliability, availability, scalability and manageability;
• Sources and consumers that may soon be upgraded or retired; and
• Overall complexity and total cost of ownership.

Composite can help to assess the current state of your information architecture using the following:

- **Data Introspection and Relationship Discovery Functionality** – The Composite Data Virtualization Platform provides automated data introspection and relationship discovery capabilities. Use these functions to understand your existing data sources.

- **Data Virtualization Architecture Advisory Service** – Composite Professional Services can help document and understand your current data architecture. Use this service to complement your data architecture skills and resources.

**Develop a plan to migrate from the current to a future state**

After defining the future-state architecture and current-state evaluation, use this information to do an architecture gap analysis and plan. The architecture gap analysis will enable you to do the following:

• Identify the technologies, skills, projects, and processes you need to move forward;

• Define a bridging strategy to accommodate current requirements as well as future needs; and

• Prioritize activities to move from current- to future-state efficiently and effectively.

In addition to the Architectural Advisory Services described above, Composite can provide technology expertise and project management assistance to add value to your planning efforts.

- **Data Virtualization Project Planning Advisory Service** – Composite Professional Services has extensive data virtualization project experience and domain knowledge. Use this service to efficiently plan a wide range of data virtualization architecture implementation projects.

**Align specific projects within overall architecture**

With the overall architecture plan in place, it is time to align specific data virtualization implementation projects to this new plan. This test of the new architecture ensures it will meet your objectives. At this point, consider expanding your list of stakeholders and projects to be included in the project scope. When aligning specific projects to the new architecture, more details will be added to the list of requirements.

**Identify and resolve issues**

As you proceed to align specific projects to the new architecture, new requirements will become known. Each new requirement will reveal issues that need to be resolved. Developing efficient
issue resolution process will be assuring to your project stakeholders, build trust, and enable you to progress faster.
In the Configuration and Set Up stage, you

- Establish your overall data virtualization environment, including development, test/QA, staging, production, back-up and fail-over;
- Set up users, access and security; and
- Integrate with other development and management tools.

**Define data virtualization environments**

The first step is to identify the development, test/QA, staging, production, back-up and fail-over environments required for your projects. Composite recommends specific best practices detailed in the following:

- **Best Practices – Strategy for Development** – This document outlines the best practices for setting up environments for development, test/QA, staging, production, back-up and fail-over. Use it to establish migration and promotion policies and procedures that achieve optimal productivity, security, and governance.

**Install Composite Data Virtualization Products**

Once you know where to install Composite data virtualization products, Composite provides a number of documents and resources to help complete this task smoothly.

- **Installation Guide** – The Composite Installation Guide is a user manual that guides pre-installation activities, installation steps, how to use the silent mode installer, and how to uninstall Composite products. Use it as a checklist for successful installation.
- **Best Practices – Production Server Configuration** – This document outlines the best practices for configuring the production server including settings for drivers, memory, SQL engine, events and logging. It also covers hardware sizing recommendations and more. Use it to fine-tune your production server configuration.
- **Getting Started Guide** – The Getting Started Guide is a user manual that describes the fundamentals for how to get up and running quickly after installation. It explains the sample resources, data, and artifacts that come with the product. It also provides a tutorial. Use this to begin ramp Composite developers and operators.

**Set up security**

Data security is critical. Composite provides a rich set of options to ensure the level of security that meets your data governance and risk policies.
Set up access to data sources and consumers

Once the security infrastructure is set up, the next step is to set up the connections between your data sources, Composite and your data consumers. Composite provides several tools to help your database administrators perform these activities.

- **Composite Platform Technical Overview White Paper** – This white paper provides a technical overview of the data source and consumer options within Composite’s Data Virtualization Platform. Use it to identify needed connections.

- **Connecting Sources and Determining Custom Source Settings Advisory Service** – Composite Professional Services has extensive experience connecting Composite to sources and consumers. Use this service to complement expertise and resources in your DBA team.

- **Composite Data Virtualization and NOSQL Data Stores** – This paper describes the primary “Not Only” SQL data sources in the market today and how to integrate them with other sources using the Composite Data Virtualization Platform. Use it to guide your NOSQL access and integration efforts.

Integrate development and system management tools and methods

The Composite Data Virtualization Platform includes a full set of development and system management tools to be integrated with your existing development and management tools. Fortunately, your technical teams will be familiar with these tools as they leverage many of the well-known data management approaches, standards, and metaphors.

Composite includes a number of APIs to simplify integration and facilitate an integration plan that evolves as data virtualization adoption grows.

- **Composite Platform Technical Overview White Paper** – This white paper provides a technical overview of the development and management tools within Composite’s Data Virtualization Platform. Use it to identify what tools you need to integrate.

- **Technology Integration Advisory Service** – Composite Professional Services has extensive experience integrating Composite with popular development and system management tools. Use this service to complement the expertise and resources of existing development and operations teams.
In the Design and Development stage, you move from organization-wide data virtualization planning and set-up activities to the detailed project-level software engineering.

**Design required views, services, caches, etc.**

Careful planning and execution at the front-end of the development cycle leads to greater agility and lower costs down the road. The key to success is to properly design the views and data services you will use. Composite provides a number of resources to help you achieve this goal.

- **Composite Data Abstraction Best Practices White Paper** – This paper outlines how to design a semantic abstraction or data services layer in support of multiple consuming applications. Use this to understand the canonical modeling objectives and levels of abstraction across the views and services you build.

- **Data Virtualization Design Advisory Service** – Composite Professional Services has extensive experience designing views and services in Composite. Use this service to complement expertise and resources in your development teams.

- **Discovery Users Guide** – This user manual describes how to use Composite Discovery to better understand the data and relationships needed to quickly build rich data models. Use Discovery to jumpstart your modeling.

**Develop**

Composite’s graphical development environment, leveraging a well-known data modeling metaphor and powerful automation, makes building and testing views easy. Manual override capabilities offer developers even more control.

Composite provides a number of resources to assist developers at this step.


- **User’s Manual** – This Composite manual describes Composite Studio data virtualization development functionality and features. Use this as a guide for your developers’ day-to-day view and data service development efforts.

- **Data Virtualization Development Advisory Service** – Composite Professional Services has extensive experience developing views and services in Composite. Use this service to complement expertise and resources in your development teams.
- **Best Practices – Establishing Source Control** – This document, currently under construction, will outline the best practices to configure and integrate data virtualization development with popular source control systems. Use it to guide source control integration.

**Test and validate**

Because data virtualization design and development cycles iterate more quickly than traditional ETL and data consolidation approaches, be prepared to update your testing processes more often. With data virtualization, you can rapidly create and iterate data integration deliverables. As such, consider ways to include end users in testing and validation earlier in your software development lifecycle.

**Tune performance**

Once users validate the data models and views, the next step is to test queries in a production environment. This includes stress testing and fine tuning the views to ensure optimal performance. Composite automatically generates optimized queries based on cost calculations; developers can also provide rules to further optimize the queries. The resources below provide best practices to tune performance.

- **Composite Data Virtualization Performance White Paper** – This white paper explains how various factors affect performance. It also describes a simple benchmarking test including various aspects of the platform performance with single and multiple heterogeneous data sources. The performance results are presented in a summary table at the end of the paper. Use it to understand key performance considerations.

- **Best Practices – Performance Tuning** – This document outlines the best practices for optimizing queries and explains ways to fine tune rule-based and cost-based optimizations. It explains execution plans and describes fundamental JOIN algorithms in detail. Use it to guide your tuning efforts.

- **Best Practices – Caching Concepts and Capabilities** – This document outlines the best practices for configuring caching options. It describes database versus file caching and table versus view-based caching as well as procedure-based caching. It provides examples and recommendations on when to use each option. Use it to guide your caching development and deployment efforts.

**Document views and services**

Composite automatically creates and maintains rich descriptive metadata for all resources built within the Composite Data Virtualization Platform.
In the Deployment stage, you move new data virtualization solutions into production.

**Identify service levels and management requirements**

Work with the end user community and your operations groups to agree on service levels. Many organizations have standard SLAs across all applications, but there may be items unique to data virtualization that you should consider. For example, data source SLAs may impact data virtualization SLAs.

**Develop high availability and disaster recovery policies and processes**

High availability and disaster recovery are standard operating processes used to ensure operational SLAs. Composite provides broad high availability and disaster recovery functionality and options for integration into your procedures.

- **Active Cluster Administration Guide** – This manual describes how Composite Active Cluster can be used to ensure high availability and disaster recovery. Use it to guide how to cluster Composite Information Server instances across your data virtualization environment to meet high availability and disaster recovery SLAs.

**Deploy completed objects on downstream environments**

In the Configuration and Set Up stage mentioned earlier, you set up the appropriate development, test/QA, staging, production, back-up and fail-over environments. At this step, deploy completed Composite resources (views, services, caches, etc.) across this landscape. Composite provides the following document to help complete this task smoothly:

- **Best Practices – Promotion Procedures and Methodology** – This document, currently under construction, will recommend the best practices and methodologies for promoting objects beyond development. Use it to guide your deployment processes.

**Document deployment processes and deployments**

Each Composite environment logs resource change activity to provide a complete deployment history.
In the Operation stage, attention shifts to how to ensure reliable, high performance operations that achieve SLA objectives.

**Ensure reliable operation**

Reliable operation is a product of many variables such as hardware, system software, applications software, management tools, procedural discipline, and more. Typically, rather than start from scratch, Composite users extend existing management tools and operational processes to include data virtualization.

Composite provides both products and professional services to help ensure reliable operations.

- **Monitor Configuration and Users Guide** – This manual describes how Composite Monitor helps to monitor the activity and health of your Composite environments. This documentation contains an overview of Composite Monitor and provides instructions for installation, configuration and operation. Use it to guide how to add Monitor to your existing system monitoring tools and processes.

- **Data Virtualization Operations Advisory Service** – Composite Professional Services has extensive experience in the operation of Composite data virtualization environments. Use this service to implement Composite Monitor and adjust and extend existing operational processes to include data virtualization.

**Administer data virtualization environments and users**

Your data virtualization environments require ongoing administration and user maintenance. Depending on the size and complexity of your data virtualization deployment and number of users, you may or may not need a full-time system administrator since many of the tasks can be automated.

- **Administration Guide** – This Composite manual describes the administration concepts and tasks for Composite Data Virtualization Platform. Use it to guide your system administrators’ ongoing administrative activities.

**Tune performance**

Performance tuning initially occurs in the Design and Development stage. However, once your data virtualization implementation is in production, you may find opportunities for further performance improvements. Composite provides several resources to assist with performance tuning during operations.

- **Best Practices – Performance Tuning** – This document outlines the best practices for optimizing queries and explains ways to fine-tune both rule-based and cost-based
optimizations. It explains execution plans and describes fundamental JOIN algorithms in detail. Use it to improve the performance of your data virtualization queries.

- **Best Practices – Caching Concepts and Capabilities** – This document describes various caching options including database versus file caching, table versus view-based caching as well as procedure-based caching. It provides examples and recommendations on when to use each option. Use it to optimize use of caching.

**Resolve problems**
Composite Support is able to resolve a range of data virtualization functionality, technology and configuration issues. Use Composite Support to select support offerings that match your service level objectives including:

- Standard email and telephone support, Monday thru Friday, 6:00am to 5:00pm PST;
- Premium email and telephone support, 24/7, 365 days per year; and
- Self-service knowledge bases and patch downloads.

**Implement patches**
Implementing patches is an ongoing process. Composite Support helps you to be proactive and avoid problems with the latest Composite maintenance releases and bug fixes. Composite Support provides complete installation instructions and can assist with Composite patch activities. See the Composite Support site for details.
In the Continuous Improvement stage, your attention shifts to optimization of your data virtualization investments.

Continuous improvement ranges from fine-tuning to major re-engineering efforts. We recommend a practical approach to continuous improvement that lets you take advantage of the latest releases and all its improvements to existing features while expanding use of new features as you see fit.

**Implement new Composite releases**

Composite delivers a major release approximately every nine months. These releases often add new features. But more importantly to existing customers, they always improve three main categories of capabilities: performance, security, and data source access.

Composite Support assists in upgrading from one release to the next by providing complete installation instructions. See the Composite Support site for details.

In addition, Composite Professional Services has upgrade knowledge and best practices.

- **Best Practices – Upgrading Composite Data Virtualization Platform Versions** – This document provides recommended procedures for upgrading your Composite products. Use it to guide administrators tasked with upgrading Composite.

**Adopt additional Composite features**

As you execute projects to create and deploy new data virtualization solutions, features that you have yet to leverage may become critical. New features from one release to the next are documented in the release notes available from Composite Support. A summary of features is listed in the various Composite product data sheets, and a comprehensive list is available in the reference manual. Ask your Composite Systems Engineer about new features.

- **Composite Product Data Sheets** – Composite product data sheets summarize key capabilities. Use these to consider how and when to implement new Composite Data Virtualization Platform features.

- **Reference Manual** – This user manual is the most comprehensive documentation on the Composite products. Use it to understand how to leverage new features in detail.

**Consider new use cases**

Improve data virtualization usage and value in your organization by learning how other companies use it. Composite shares general usage patterns and specific project use cases developed by our customers.
• **Composite Data Virtualization Usage Pattern White Paper** – This white paper introduces the most popular data virtualization usage patterns, the business and IT challenges behind them and the solutions Composite delivers to overcome these challenges. Use it to help you identify areas to take advantage of data virtualization.

• **Composite Customer Case Studies** – These case studies provide detailed insight into Composite data virtualization use cases. Use them to identify specific data virtualization opportunities.

**Develop reuse strategies**

Reuse is a major value proposition of data virtualization. Once you have implemented one or more deployments, consider how Composite resources (views, data services, caches, etc.) can be leveraged to save time and effort on future projects. Composite provides a number of resources to help you achieve this goal.

• **Composite Data Abstraction Best Practices White Paper** – This white paper outlines how to design a semantic abstraction or data services layer in support of multiple consuming applications. Use this to clarify your approach to reusability across the views and services you build.

• **Data Virtualization Design Advisory Service** – Composite Professional Services has extensive experience designing reusable views and services in Composite as well as extending existing ones for additional usage. Use this service to complement expertise and resources in your development teams.

**Improve run-time operations**

Improving operations is just as important as improving functions and expanding use cases. From the end users’ perspective, a highly available running environment provides the confidence to endorse continued data virtualization expansion. Composite provides resources to help fine-tune operations as well as a guide for setting up centralized centers of excellence such as an Integration Competency Center (ICC).

• **Case Study: Pfizer Delivers Effective Data Integration Capabilities via a Shared Services Model** – This research report (ID Number: G00159896) is available to Gartner Research subscribers and describes data integration and delivery operations led by the shared services team at Pfizer, a Composite Software customer. Use it to understand the best practices and critical competencies needed to reduce costs and improve the operational execution of data virtualization.

• **Best Practices – Performance Tuning** – This document outlines the best practices for optimizing queries and how to fine-tune both rule-based and cost-based optimizations. It explains execution plans and describes fundamental JOIN algorithms in detail. Use it to improve the performance of your data virtualization queries.
CONCLUSION

Adoption of the Composite Implementation Methodology and Best Practices can significantly accelerate your data virtualization success. Based on Composite’s unmatched industry experience and expertise, join hundreds of who have leveraged these best practices, white papers, reference manuals, advisory services and more from Composite. Easy to blend with your own methods and best practices, our end-to-end methodology and proven best practices demonstrate why Composite Software is the data virtualization gold standard.
ABOUT COMPOSITE SOFTWARE

Composite Software, Inc. is the data virtualization gold standard at ten of the top 20 banks, six of the top ten pharmaceutical companies, four of the top five energy firms, major media and technology organizations; and multiple government agencies. These are among the hundreds of global organizations with disparate, complex information environments that count on Composite to increase their data agility, cut costs and reduce risk. Backed by nearly a decade of pioneering R&D, Composite is the data virtualization performance leader, scaling from project to enterprise for data federation, data warehouse extension, data virtualization layer and cloud data integration. Composite Software is a privately held, Silicon Valley-based corporation. For more information, please visit http://www.compositesw.com. Or follow Composite on http://twitter.com/compositesw.

Composite Software and Data Virtualization Leadership Series are registered trademarks of Composite Software, Inc. Copyright © Composite Software, Inc. 2011.