Eight Ways Composite Data Virtualization Adds Value to Enterprise Data Warehousing

COMPOSITE
SOFTWARE

TABLE OF CONTENTS

MAXIMIZING VALUE FROM ENTERPRISE DATA WAREHOUSE INVESTMENTS

Supporting critical, yet ever changing information requirements in an environment of ever increasing data volumes and complexity is a challenge well understood by large enterprises and government agencies today.

This inexorable pressure has and will continue to drive the demand for enterprise data warehouse centric solutions, as an array of business intelligence, predictive analytics, data and content mining, portals, and other key applications rely on data sourced from enterprise data warehouses.

However, business change often outpaces enterprise data warehouse evolution. And while useful for physically consolidating and transforming a large portion of enterprise data, significant volumes of enterprise data continues to reside outside the confines of the enterprise data warehouse. Further, enterprise data warehouses themselves require support throughout their lifecycle, driving demand for solutions that prototype, migrate, extend, federate and leverage enterprise data warehouse assets.

Composite data virtualization complements enterprise data warehouses by providing a range of flexible data integration techniques that let you preserve and extend existing enterprise data warehouse investments.

In this paper you will learn 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.

DATA WAREHOUSE EXTENSION

Problem – Users Require Data From Outside the Data Warehouse

Enterprises and government agencies overwhelmed by scattered data silos and exponentially growing data volumes have deployed data warehouses to meet many of their reporting requirements. However a number of sources remain outside the warehouse. Providing users with complete business insight so they can achieve revenue, cost and risk management goals requires both:

- Historical data from the warehouse and up-to-the-minute data from transaction systems or operational data stores
- Summarized data from the warehouse and drill-down detail from transaction systems or operational data stores
- Internal data from the warehouse and external data from outside sources.

Solution – Use Data Virtualization to Extend Existing Data Warehouses

You can use Composite data virtualization to federate data warehouse data with additional sources, in effect extending existing data warehouse schemas and data. These complementary views let you add current data to historical warehouse data, detailed data to summarized warehouse data, external data to internal warehouse data, and more to quickly and easily work around the fact that key data your users need resides outside your consolidated data warehouse repositories.

In the integration pattern shown in Figure One below, the Composite Information Server hosts new complementary views that integrate additional RDBMS and Web Service data sources as well extend an existing source, Packaged Applications. (See Figure One.)

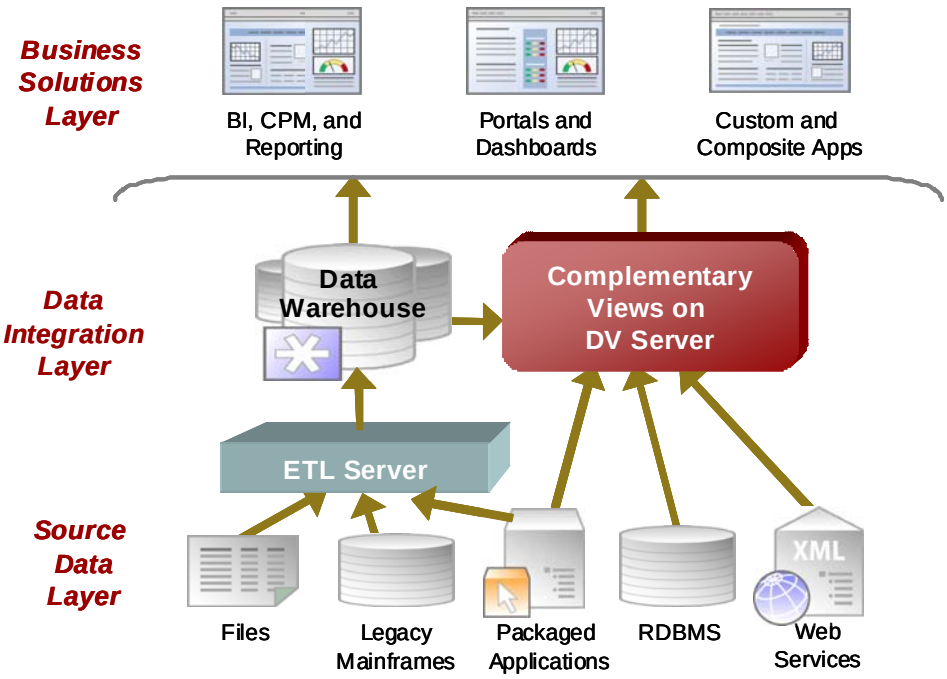

Figure One: Use Data Virtualization to Extend Existing Data Warehouse Schemas and Data

Selected Examples

Combining Up-to-the-minute and Historical Data – To optimize deployment of repair crews and equipment across more than 10,000 production oil wells, an energy company uses Composite data virtualization to federate real-time crew, equipment, and well status data from the wells and SAP's maintenance management system with historical surface, subsurface, and business data from the enterprise data warehouse. The net result is faster repairs which translate into more uptime and more revenue.

Combining Internal and External Data – To analyze pharmaceutical sales and marketing effectiveness, this life sciences company uses Composite data virtualization to federate internal prescription sales data from their sales data warehouse with externally sourced competitor sales data from by a industry data services provider. This total market view enables more effective sales and marketing programs resulting in additional revenues.

360° VIEW EXTENSION OF MASTER DATA MANAGEMENT HUB

Problem – Key Detail Data Lives Outside the MDM Hub

As data silos have proliferated, the business case for improving control and leveraging your master data has become compelling.

- **Customer Master Data**. Grow revenue by selling additional offerings to existing customers.
- **Product Master Data.** Increase supply chain efficiency by eliminating duplicate products.
- **Employee Master Data.** Improve employee productivity and retention by unifying personnel information.

Multiple MDM vendors have responded to this demand. However, their applications alone cannot fully support all your requirements. Complementary data integration solutions are needed to deal with related data such as order histories, inventory balances, and payroll records maintained outside your MDM hubs across a myriad of complex, disparate data silos.

Solution – Use Data Virtualization to Extend the Master Data and Provide a Complete 360° View

MDM hubs maintain and control critical master data attributes, but not the detailed transaction histories and other related data maintained and controlled in dozens of other systems across your extended enterprise. Composite data virtualization leverages master data from your hub as the foreign key to quickly and easily federate your master data with additional transactional and historical data so you can get a complete single view of your customer, product, and employee information.

In the integration pattern shown in Figure Two below, the Composite Information Server hosts new complementary views that integrate additional RDBMS and Web Service data sources as well extend an existing source, Packaged Applications. (See Figure Two.)

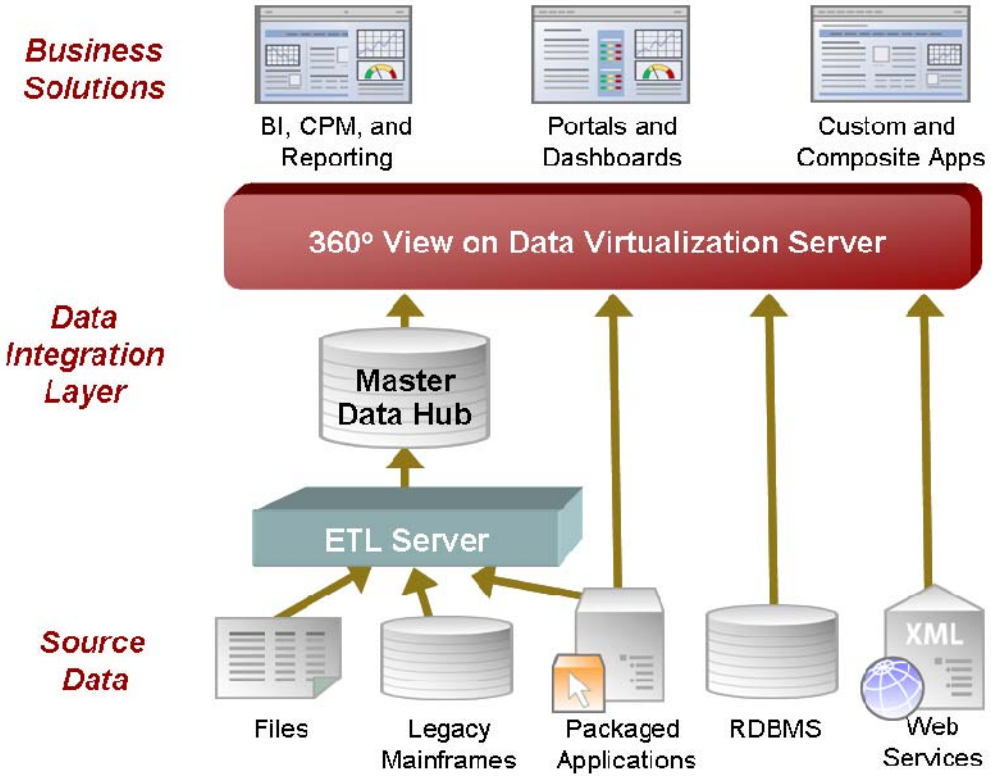

Figure Two: Use Data Virtualization to Create 360° View of Parties and Products

Selected Examples

Providing a 360° View of Customer Data – To maximize revenue per customer and customer service levels, this European mobile phone operator uses Composite data virtualization to combine customer service data from their customer reporting data warehouse, billing data from their financial systems, and call and configuration data from their operational support systems to deliver a 360° view of customers to their customer service representatives. Faster issue resolution and more productive up-sell programs reduce churn and increase revenues.

Providing a 360° View of Employee Compensation - To ensure retention, this money center bank uses Composite to federate employee master data with myriad internal benefits and compensation systems as well external payroll services to provide the employees with a 360° view of their total compensation through a self-service employee benefits portal. Securely exposing this information improves retention and lessens HR staff workload.

DATA WAREHOUSE FEDERATION

Problem – Multiple Data Warehouses Means Even More Data Silos

One of the main reasons enterprises implement data warehouses is to overcome the various transaction and analytic system silos typical in most large enterprise and government agencies today. However, for a number of often pragmatic reasons, the single "enterprise" data warehouse remains elusive. Instead, for these same reasons, multiple data warehouses and data marts have been developed and deployed, in effect perpetuating rather than overcoming the data silo issue.

Optimizing business performance requires data from across these various warehouses and marts. But physically combining multiple marts and warehouses into a singular and complete enterprise-wide data warehouse is often too costly and time consuming.

Solution – Use Data Virtualization to Federate Multiple Data Warehouses

You can use Composite data virtualization to federate multiple physical warehouses, for example to combine data from the sales and financial warehouses or to combine two sales data warehouses after a merger. This approach achieves logical consolidation of warehouses by creating an integrated view across them, using abstraction to rationalize the different schema designs.

In Figure Three below, the Composite Information Server hosts federated warehouse views that logically integrate both data warehouses. (See Figure Three)

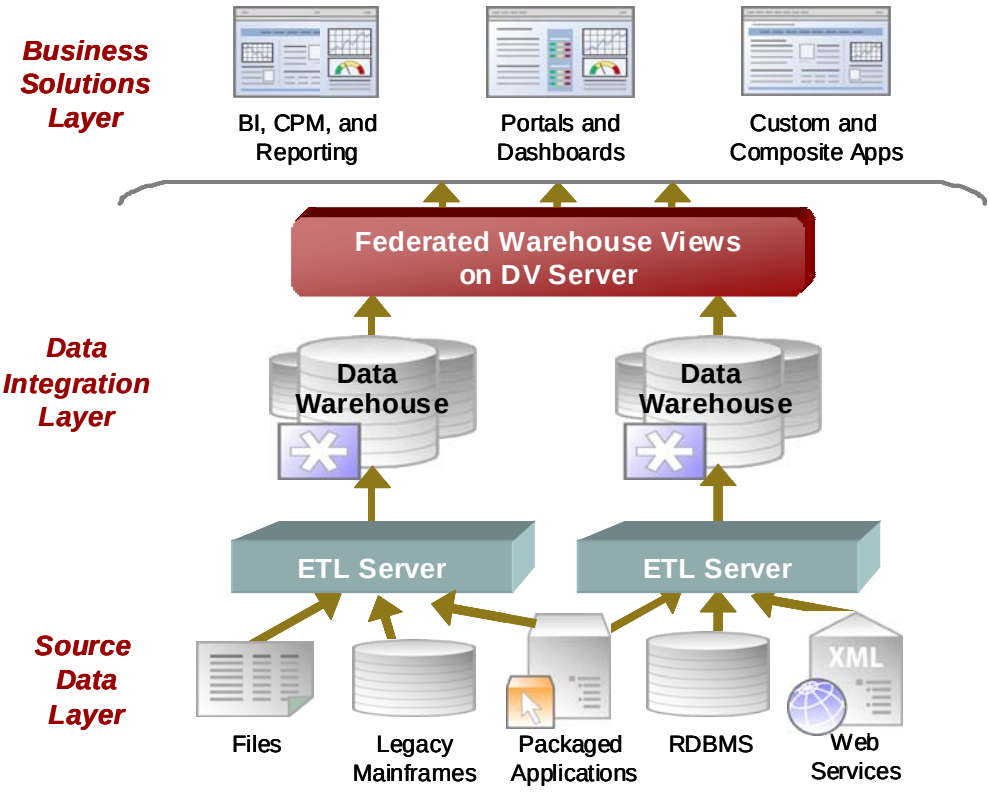

Figure Three: Use Data Virtualization to Federate Multiple Data Warehouses or Marts

Selected Examples

Federating Financial Trading Data Warehouses – To enable more flexible customer self-service reporting and meet SEC compliance reporting mandates, this prime brokerage uses Composite data virtualization to federate equity, fixed income, and other investment positions and trades information from siloed trading data warehouses. The net result is higher customer satisfaction and lower reporting costs.

Federating Research and Development Data Warehouses – To enable research scientists to access and analyze data from research, clinical trial, FDA submission, and other data warehouses, this pharmaceutical company uses Composite data virtualization to federate these diverse warehouse sources. Scientists use this federated data to accelerate time to market for new compounds and drugs, thereby increasing revenues.

DATA WAREHOUSE HUB AND VIRTUAL DATA MART SPOKE

Problem – Data Mart Proliferation Is Costly and Degrades Data Quality

A typical data warehouse pattern is a central data warehouse hub with satellite data marts as spokes around the hub. These marts typically use a subset of the warehouse data and are used by a subset of the data warehouse users. Sometimes these marts are created because the analytic tools used require data in a different form than the warehouse. However, sometimes they are created to get around the controls provided by the warehouse, "rogue data marts" so to speak. Regardless of the reason, every additional mart adds cost and compromises data quality.

Solution – Use Data Virtualization to Create Virtual Data Marts

You can use Composite data virtualization to provide virtual data marts that eliminate, or at least significantly reduce, the need for physical data marts around your data warehouse hubs. This approach uses abstraction to transform the warehouse data to meet specific consuming tool requirements and user query requirements, while still preserving the quality and controls inherent in the data warehouse.

In the integration pattern shown in Figure Four below, the Composite Information Server hosts virtual data marts that logically abstract and serve specific analytical reporting requirements. (See Figure Four.)

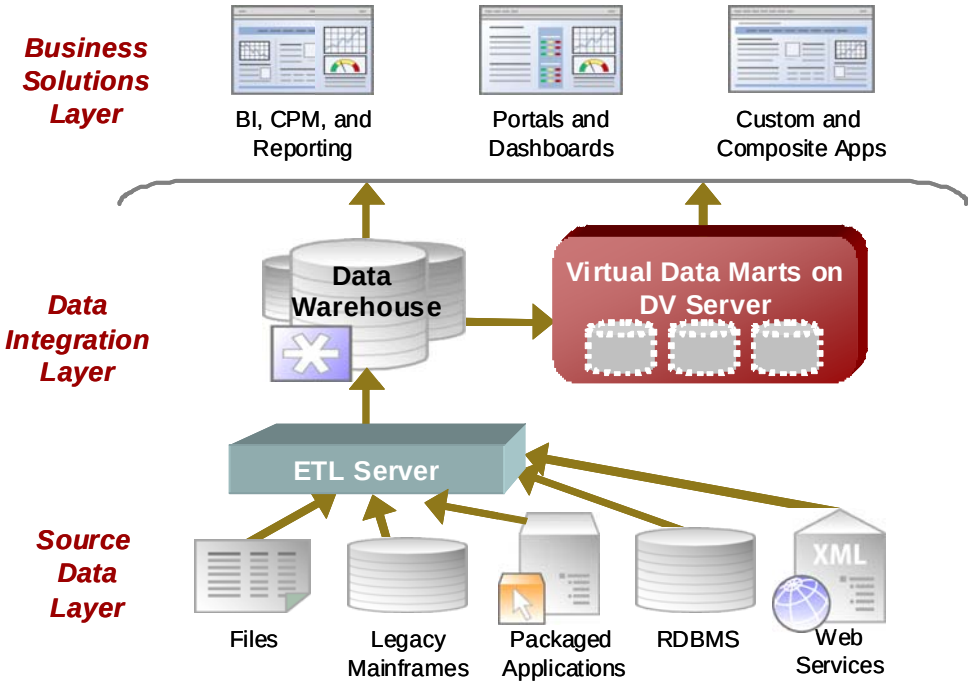

Figure Four: Use Data Virtualization to Virtualize Spoke Data Marts

Selected Examples

Eliminating Rogue Data Marts –A mutual fund company uses data virtualization to enable more than 150 financial analysts to build portfolio analysis models with MATLAB® and other analysis tools leveraging a wide range of equity financial data from a 10 terabyte financial research data warehouse. Prior to introducing data virtualization, analysts frequently spawned new satellite data marts with useful data subsets for every new project. To accelerate and simplify data access and to stop the proliferation of costly, unnecessary physical marts, the firm instead used data virtualization to create virtual data marts formed from a set of robust, reusable views that directly accessed the financial warehouse on demand. This enables analysts to spend more time on analysis and less on access, thereby improving portfolio returns. The IT team has also eliminated extra, unneeded marts and all the costs that go with maintaining them.

Supporting Diverse Analytics with Virtual Marts – To provide oil well platform data from a central Netezza data warehouse to engineers, maintenance managers, and business analysts each requiring different slices of the data, optimally formatted for their wide range of specialized analysis tools including Business Objects, Excel, Tibco Spotfire, Matrikon ProcessNet, Microsoft Reporting and more, this energy company uses Composite data virtualization. Composite's ability to build virtual views and services quickly enabled rapid response to new ad hoc queries. Rapid time to data, combined with ease of abstraction (convert from warehouse-stored format to tool-required format), and lower costs encourages analysts to leverage the warehouse as the single source of truth rather that replicate data in local, rogue data marts.

INTEGRATING DATA WAREHOUSES INTO ENTERPRISE INFORMATION ARCHITECTURES

Problem – Advanced Enterprise Information Architectures Must Include All Information Assets

While the enterprise data warehouse is often the primary source for significant volumes of enterprise information, other sources are also critical today. This has the potential to increase in the future, as data grows exponentially and complexity continues unabashed. Increasingly enterprises are seeking unified ways to integrate warehouse and other data in an enterprise-wide information architecture. According to Forrester Research, "new architectural approaches such as information-as-a-service (IaaS) have emerged to provide flexible, real-time, service-oriented data integration and data-quality capabilities that support both structured data and unstructured content, delivering a true information integration platform." (1)

Solution – Integrate Data Warehouses into Enterprise Information Architectures

Composite Data virtualization integrates data warehouses into an unified enterprise information architecture. The data virtualization middleware forms an enterprise data virtualization layer that is home to a logical schema covering multiple consolidated and virtual sources in a consistent and complete fashion. In design, developers use data virtualization design tools to develop these semantic abstractions in the form of web services or relational views. At run time, end user-level applications, reports or mash-ups can call these web data services on demand to query, federate, abstract and deliver the requested data to these information consumers. (See Figure Five.)

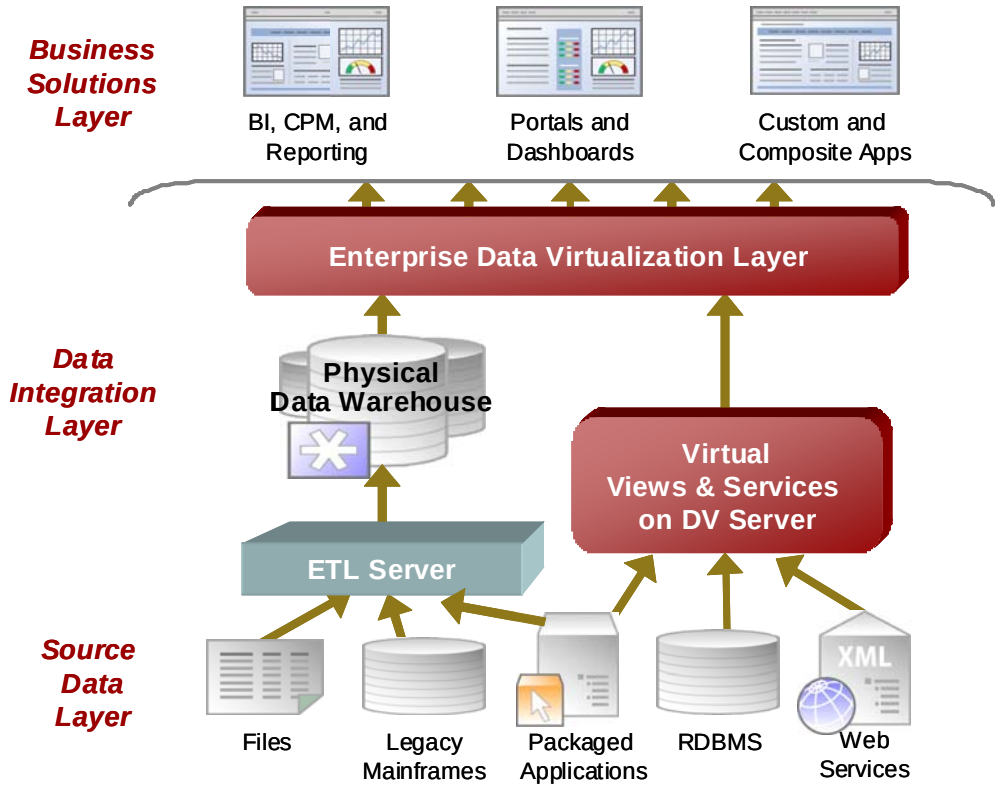

Figure Five: Integrating Data Warehouses into an Enterprise Data Virtualization Architecture

Selected Examples

Virtualizing Refinery Data Enterprise-wide – To provide disparate data warehouse and operational data from more than a dozen refineries to diverse technical and business user communities globally, this energy company deployed Composite data virtualization on an enterprise scale. This common approach allowed them to increase refinery yields, proactively maintain equipment, and comply with a myriad of regulations more consistently for less.

Sharing Intelligence Data across Government Agency Boundaries – To enable intelligence analysts to use information from other agencies and better control threats, multiple government agencies leverage a common Composite data virtualization layer. This allows other agencies such as the Drug Enforcement Administration and the Immigration and Naturalization Service to access passenger, crew and manifest data from a U.S. Coast Guard port arrivals data warehouse, for example.

1. **2009 Update: Evaluating Integration Alternatives**, *Scenario-Based Guidance For Choosing Products That Provide Application, Process, And Data Integration Features*, Copyright (r) 2009, Forrester Research, Inc. Ken Vollmer, Rob Karel, Larry Fulton, Noel Yuhanna. with Gene Leganza and Matt Czarnecki

COMPLEMENTING THE ETL PROCESS

Problem – ETL Alone Is Not Always the Most Effective Way to Load a Warehouse

Extract, Transform, and Load (ETL) middleware is the tool of choice for loading data warehouses. However, there are some cases where the ETL tools are not the most effective approach, for example where:

- ETL tools lack interfaces to easily access source data, for example data from packaged applications such as SAP or new technologies such as Web services
- Readily available, existing virtual views or data services can be reused rather than building new ETL scripts from scratch
- Tight batch windows require access, abstraction, and federation activities to be pre-processed and virtually staged in advance of ETL processes.

Solution – Use Data Virtualization to Preprocess Data for ETL

You can use Composite data virtualization to complement your ETL tools to gain greater flexibility when loading your data warehouse. Your ETL tools can leverage virtual views and data services as inputs to their batch process, appearing as simply another data source. This integration pattern also lets you integrate data source types that your ETL tool cannot easily access as well as reuse existing views and services, saving time and costs. Further these abstractions do not require your ETL developers to understand the structure of, or interact directly with, your actual data sources, significantly simplifying their work and reducing time to solution.

In the integration pattern shown in Figure Six below, the Composite Information Server complements ETL by providing access, abstraction and federation of packaged applications and Web services data sources. (See Figure Six.)

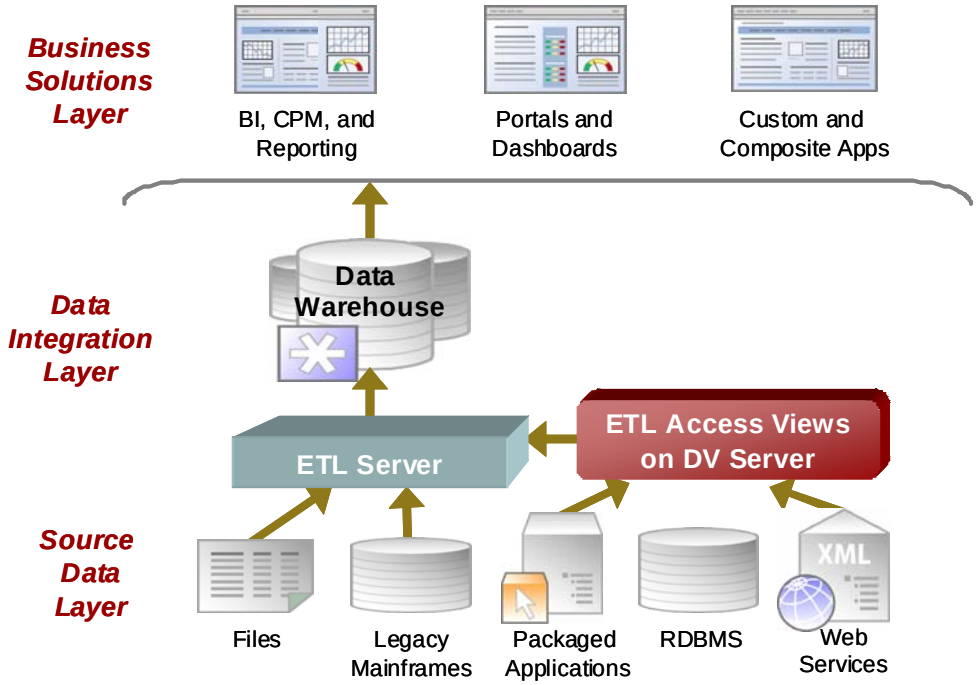

Figure Six: Using Virtual Views to Simply Packaged Application and Web Service Access, Abstraction, and Federation

Selected Examples

Preprocess SAP Data – To provide the SAP financial data required for their financial data warehouse, this energy company uses Composite data virtualization to access and abstract SAP R/3 FICO data. Composite replaced an error-prone, SAP data expert intensive, flat file extraction process that would not scale across their complex SAP landscape. The results include more complete and timely data in the financial data warehouse enabling better performance management.

DATA WAREHOUSE PROTOTYPING

Problem – Data Warehouse Development Takes Too Long

Everyone understands that building a new data warehouse from scratch is a large undertaking that requires significant design, development, and deployment efforts. One of the biggest issues is the level of effort required to affect a schema change, a frequent activity early in a warehouse's lifecycle. This change process requires modification of both the ETL scripts and physical data in the warehouse and thus becomes a bottleneck that slows new warehouse deployments. This problem does not go away later in the lifecycle; it just lessens as the pace of change slows.

Solution – Use Data Virtualization to Rapidly Prototype and Quickly Meet New Requirements

You can use Composite data virtualization to rapidly prototype and quickly meet new requirements in an early stage of a new data warehouse initiative or later as you add new data sources, federate data in different ways, and or meet new reporting needs.

In this integration pattern, the Composite Information Server serves as the prototype development environment for a new data warehouse shown in Figure Seven. At this prototype stage, you build a virtual data warehouse rather than a physical one, saving the time to build the physical warehouse. This virtual warehouse includes a full schema that is easy to rapidly iterate as was as a complete functional testing environment. (See Figures Seven and Eight)

Once the actual warehouse is deployed, the views and data services built during the prototype stage still have value for prototyping and testing subsequent warehouse schema changes that arise as business needs or underlying data sources change.

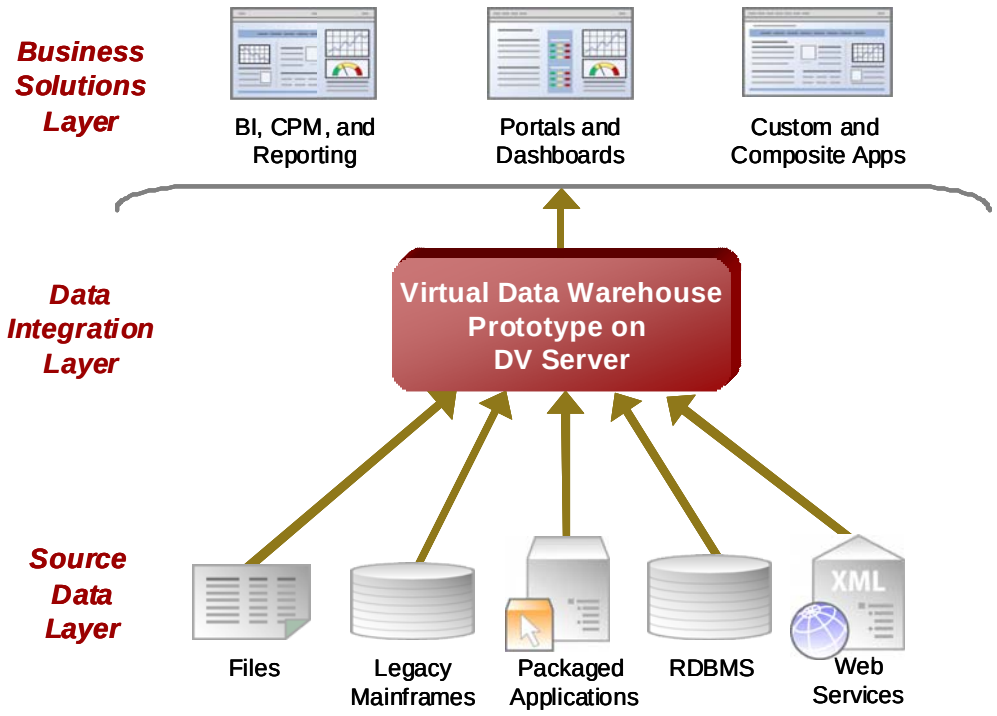

Figure Seven: Virtual Data Warehouse Serves as Prototype to Enable Rapid Development

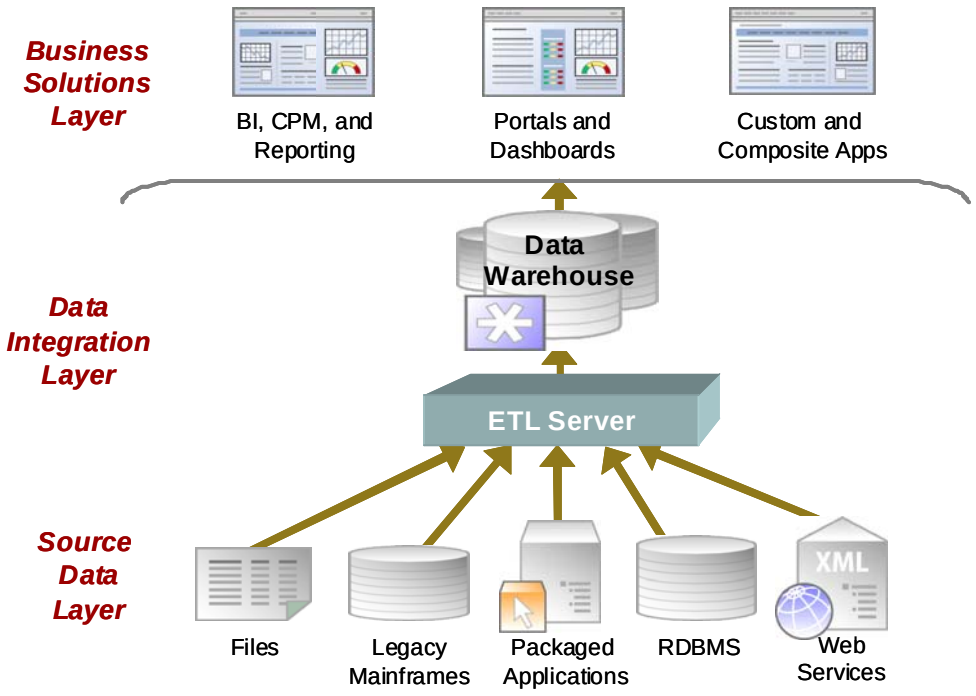

Figure Eight: Prototype Virtual Data Warehouse Replaced by Actual Data Warehouse

Selected Examples

Prototyping New Data Warehouses – To reduce time to solution for new data warehouse projects and changes to existing ones, this government agency uses Composite data virtualization. Time spent getting the data right has proven to be four times faster than a directly building the ETL and warehouse, even when including the subsequent translation of Composite views into ETL scripts and physical warehouse schemas.

DATA WAREHOUSE MIGRATION

Problem – Moving To a New Data Warehouse Risks Reporting Continuity

There are a number of reasons to migrate a data warehouse. One is cost savings. Many enterprises are finding that data warehouse appliances can significantly reduce data warehouse total cost of ownership. Another is mergers and acquisitions. In this case, duplicate financial data warehouses need to be rationalized. A third is standardization. Here an enterprise or government agency may want to rationalize various warehouses based on disparate warehouse technology platforms by moving to a standard platform.

Regardless of the reason for the migration, in every case the reporting and analysis supported by the migrating data warehouse must continue to run seamlessly.

Solution – Use Data Virtualization to Insulate Reporting Users during Data Warehouse Migrations

You can use Composite data virtualization to insulate reporting users from the impact of data warehouse migrations. Composite data virtualization removes reporting risk by inserting a virtual reporting layer between your warehouse and your reporting systems. Decoupling these systems enables the reporting to continue before, during and after the migration.

The integration pattern shown in Figure Nine below depicts the original state prior to migration. (See Figure Nine.)

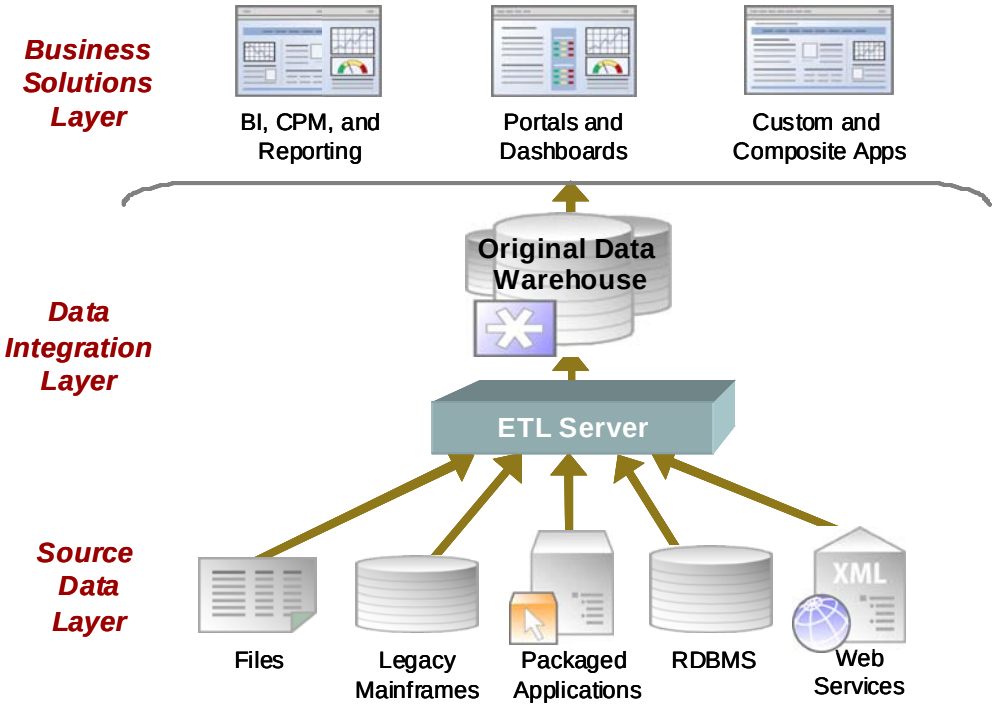

Figure Nine: Original Data Warehouse Prior to Migration

In the integration pattern shown in Figure Nine, the Composite Information Server is implemented between the reports and the warehouse. It is used to build and host the reporting data virtualization layer, a new set of views and data services that the reporting systems can use to query their data. While this requires the extra effort to create these views, this investment not only reduces immediate migration risk, it pays off by providing long run flexibility.

You must also modify the reports to query the virtualization layer, rather than the warehouse directly. But the reporting queries would need to be rewritten anyway to query from the new data warehouse. The benefit of this approach is that it allows for a controlled migration of reports in advance of the data warehouse migration. (See Figure Ten.)

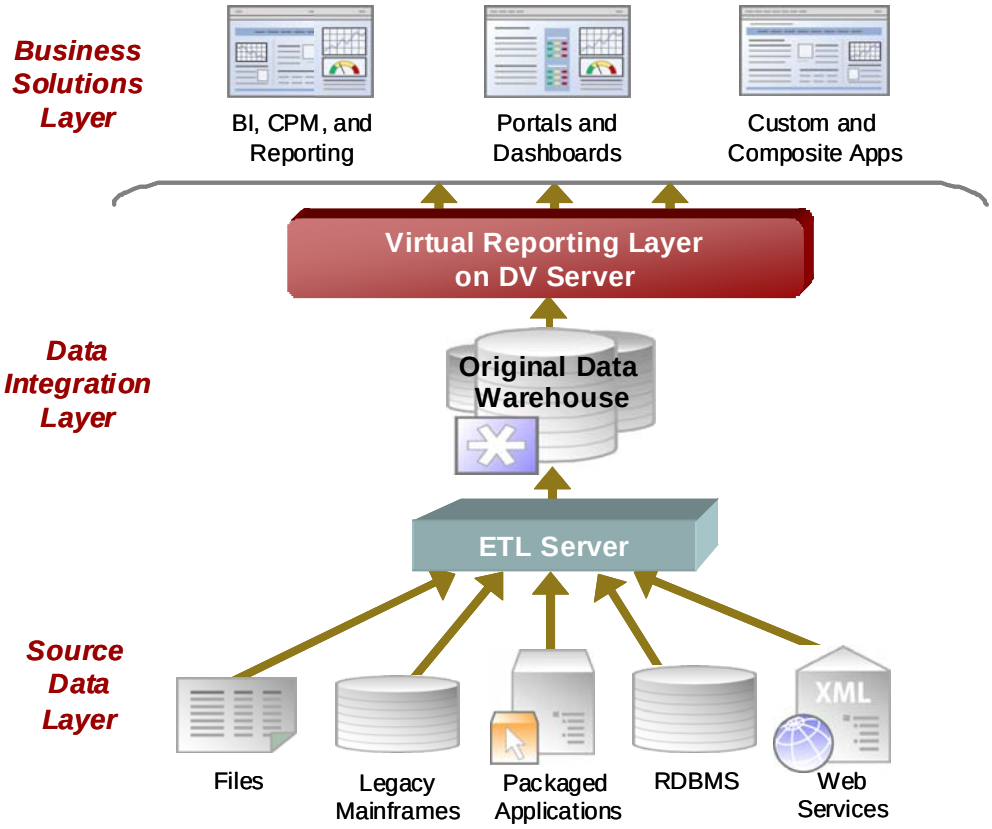

Figure Ten: Virtual Reporting Layer Added to Decouple Reports from Original Warehouse

The integration pattern shown in Figure Eleven shows the next step in the process where the new warehouse and supporting ETL are brought on line and the old warehouse is retired. Composite data virtualization insulates your reporting users at this step by enabling a controlled migration of the views in the reporting layer. Each existing view can be cloned, modified to point at the new warehouse, and tested before the actual cutover, thereby insulating the reporting users from undo risk. Further, the virtual reporting layer is easily extensible for adding more sources or supporting new reporting solutions. (See Figure Eleven.)

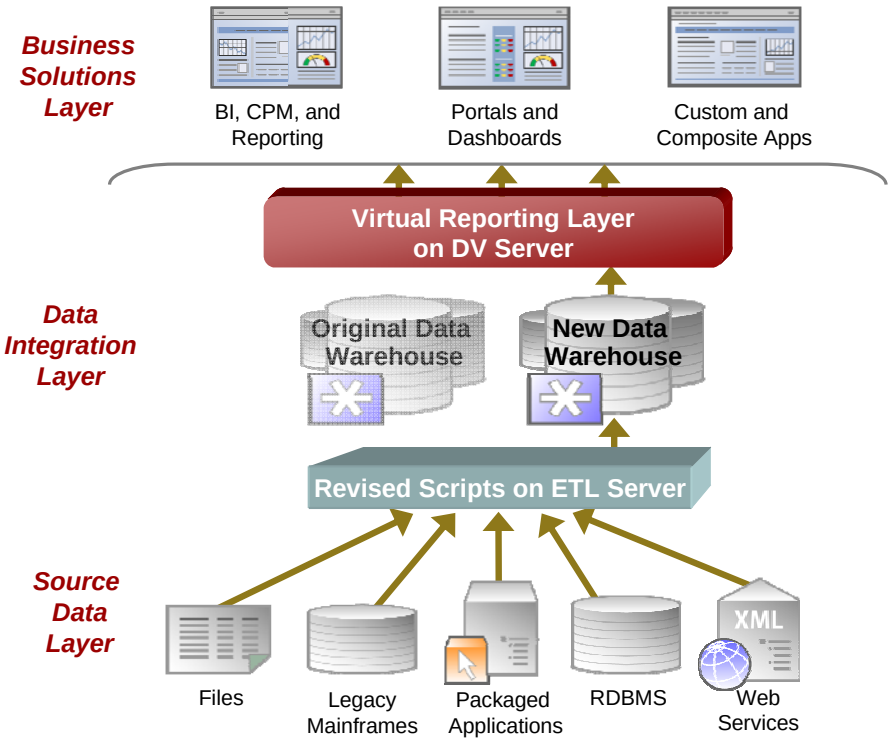

Figure Eleven: Virtual Reporting Layer Serving Data from New Data Warehouse

Selected Examples

Data Warehouse Appliance Migration – To reduce data warehousing total cost of ownership by migrating to a data warehouse appliance, this large technology company used Composite data virtualization to decouple their reporting from their data warehouse. Data warehousing cost reductions were achieved while reporting was successfully migrated without interruption.

ABOUT COMPOSITE SOFTWARE

Composite Software, Inc. is the leading independent provider of data virtualization software. Global organizations including 10 of the top 20 banks, five of the top ten pharmaceuticals, leading energy, media, and technology companies along with U.S. Defense and Intelligence agencies, use Composite's technology to integrate disparate data--regardless of location or source format--and fulfill critical information needs, faster for less. Composite's data virtualization platform scales from individual business applications to enterprise-wide Information-as-a-Service architectures, automating the entire life cycle, while complementing traditional data warehousing investments. Founded in 2002, Composite Software is a privately held, venture-funded corporation based in Silicon Valley. For more information, please visit www.compositesw.com

Headquarters
2655 Campus Drive
Suite 200
San Mateo, CA 94403
Phone: 650-227-8293
Fax : 650-227-8199
info@compositesw.com