



Data Virtualization: Extending data integration capabilities beyond the data warehouse



DATA VIRTUALIZATION: Extending data integration capabilities beyond the data warehouse

Businesses in every industry have seen extraordinary growth in data volumes in recent years, with no end in sight. This growth has been a result of many factors, such as an unprecedented proliferation of applications, the rapid growth of cloud and mobile computing, social media, the Internet of Things and more. Each of these systems generates streams of data that more often than not reside in separate silos across the business. Data integration is the process used to provide for unified views and access to these disparate data sources. With strong data integration capabilities in place this great ocean of data can become an asset to the enterprise. It is transformed into improved insight into the business through enhanced reporting and analytic capabilities; and, new data driven applications that increase customer service levels and create operating efficiencies.

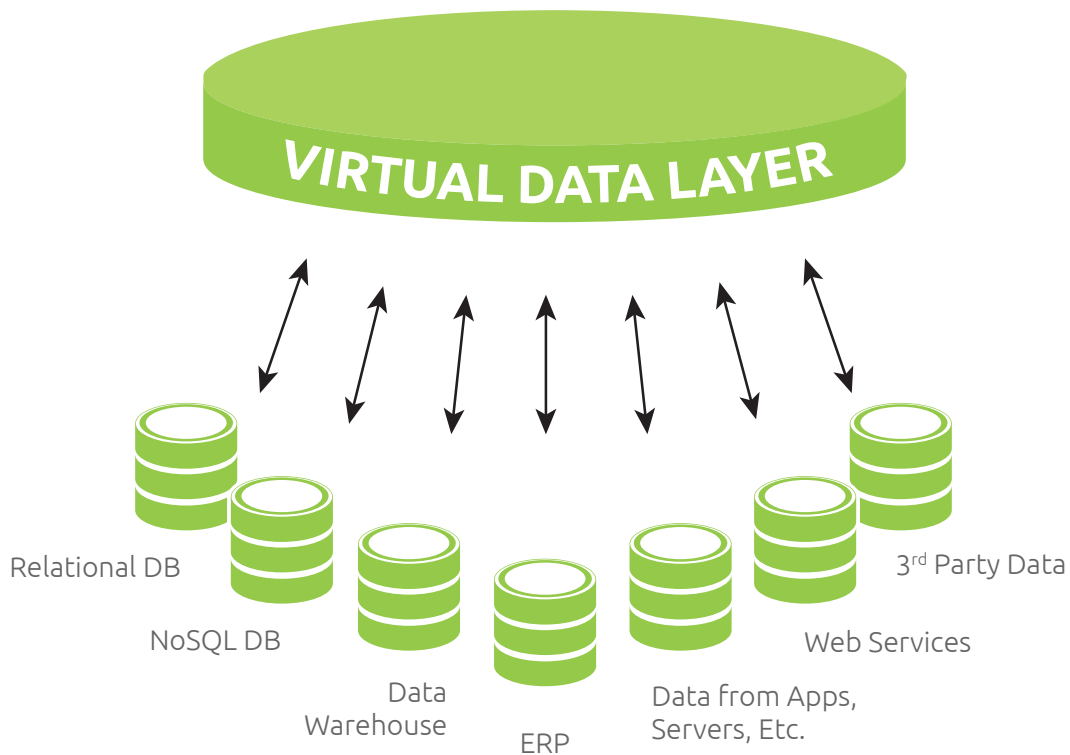
The traditional approach to data integration has been to invest in a data warehouse. This is a central repository of data integrated from multiple sources. It works by taking otherwise siloed data from sources such as marketing, sales, ERP, the supply chain, 3rd party applications, etc. and brings them together in the data warehouse. This is typically a time- and resource-intensive process to set up and maintain, since data is essentially replicated from its source to the data warehouse using “Extract, Transform, Load” (ETL) technology. Once in place, the data warehouse is typically used as the prime data source for reporting and analytics applications across the enterprise.

The challenge for many businesses is that even with a data warehouse and supporting tools in place, there are still limitations in their ability to deliver “on demand” reports and dashboards as these typically require data integrated from multiple disparate sources, not all of which are contained in the data warehouse. Happily, advances are being made to alleviate this challenge while retaining—and even enhancing—the return on the investment made in the data warehouse.

At the forefront of this improvement is a rapidly emerging approach to integration called Data Virtualization. This technology abstracts data from multiple sources creating a unified virtual data layer that provides even end-users easy access to the underlying source data. The diagram below illustrates how it works:

FIGURE 1

Graphic depiction of how data virtualization works.



With Data Virtualization there is no need for the time consuming process of preparing and manipulating source data. A new data source can typically be integrated in hours. The virtual data layer effectively becomes a large database of all of the underlying source data.

THE IMPACT OF DATA VIRTUALIZATION

Data Virtualization is a strong complement to a company's data warehouse in that it expands its data integration capabilities and preserves the existing investment of time and technology. To be sure, a production ready data warehouse system is of great value in meeting a company's standard reporting and analytics needs. Where Data Virtualization creates significant incremental value for a company with an existing data warehouse is as follows:

Versatile Reporting & Analytics Tool:

Data Virtualization allows multiple data sources to be included in a virtual data layer with little time or cost to set up and maintain. This means data sources such as transactional data, log data, application usage data, operational metrics, or IoT data can now be part of a company's virtual database. Data Virtualization delivers important new data sets in **hours or days** (rather than weeks

or months) for special-purpose dashboards and analytics, or business critical reporting. This allows you to be highly agile in responding to the needs of individuals or departments, delivering insight across the enterprise. A few practical examples are as follows:

- A sales executive feels there is great value in quickly overlaying key information stored in separate silos to understand why the company's sales seem to have changed this quarter. The necessary data is stored in its sales database, ERP database, loyalty program application, and social media application. A virtual data layer is rapidly created integrating the required data sources. Dashboards, which highlight key relationships or analysis to gain deeper insight, can be quickly created to deliver the required understanding; action can then be taken to eliminate the problem.
- A company introduces two new applications for use by its employees in the same month. The IT department suddenly receives a high volume of complaints about many existing applications experiencing poor performance. Using Data Virtualization, the company has all of its application servers populate their individual operational metrics to the virtual data layer. They can now easily build dashboards and run analysis to gain insight into load, database usage, and cluster utilization from a company-wide perspective to help identify bottlenecks and problems. This leads to a quick resolution and its applications returning to expected performance levels.

Minimize Data Marts:

A data mart is essentially a subset of data from the data warehouse. It is used to deliver unique information required by a department for their reporting and analytics purposes. A data mart takes time and resources to set up, load with the required data, and operate by definition limiting the agility with which the business can get the information it needs. With Data Virtualization the required subset of data can easily be made available to the required department with no need to replicate existing data. This saves time and resources and lets you respond more quickly to changing circumstances.

Improved Application Development Capabilities.

The improved data integration offered by Data Virtualization also translates into big gains for application development. Using traditional approaches, when building an application, developers need to write specific code to integrate an application to each of its data sources. This is time consuming and introduces the complexity of dealing with multiple back end systems. Using Data Virtualization an application only needs to be integrated with one access point, the virtual data layer. This greatly simplifies the task of building applications leading to substantial time and cost savings. Data Virtualization also improves the security of an application as all underlying source data is accessed through a single point, the virtual data layer, which can be tightly monitored and controlled.

Real time data:

Data Virtualization works by accessing source data directly. This means reporting and analytics or applications built upon a virtual data layer have access to data in real time with all of the attendant benefits. This fast access is not possible with data warehousing since data is loaded into the data warehouse through regular batch processing and is therefore never fully current.

Significant Savings:

With a virtual data layer in place it is estimated a company can save as much 50% or more on its data integration costs compared to traditional approaches such as data warehousing alone¹. With Data Virtualization, less development time is required to prepare disparate data for integration and the requirements to replicate data are dramatically reduced. Finally, there is less need to maintain and manage replicated data. The savings alone can justify a Data Virtualization implementation without even taking into account the other strategic benefits mentioned above.

A “Single Point of Truth” is Possible:

With Data Virtualization a unified data model can be created for the entire enterprise while allowing the underlying disparate data sources to stay in their native state. This is a great benefit as it allows a company to have a single master view of their data, helping a company improve its data consistency. A company may have several versions of customer or supplier data stored in many data silos each containing minor inconsistencies. This leads to the obvious problems of billing/payment errors, poor levels of customer service, etc. With a unified data model most of these issues are resolved.

SUMMARY

Data Virtualization is an emerging data integration technology that has a wide range of impactful applications inside of an IT department. It is a great complement to a company's existing data warehouse solutions: it can act as a crucial tool greatly improving the speed and agility of an organization's ability to integrate and access its data. This translates into improved reporting and analytics capabilities allowing it to be much more responsive to individual and departmental needs. Data Virtualization can be used to easily integrate new data sources into existing applications. It can streamline a company's application development capabilities and offers a host of other benefits, from real time access to data to improved data consistency. Importantly, Data Virtualization can create significant savings in both time and money on data integration. It is a data integration tool that should be considered for the arsenal of every company that wants to be a data driven enterprise.

Accur8 Software

Accur8 Software is a leading data unification company, focused on high performance, scalable and accessibly priced integration technology and tools. We recognize that companies' application environments are growing in complexity as they deploy more and more software applications to drive their businesses forward. This complexity hampers business performance because valuable data from across the company is not readily available to business users or systems. It forces IT staff to waste significant time and money dealing with the never-ending cycle of trying to integrate and share needed data across the organization.

The Accur8 Integration Engine is a data unification tool designed to help companies address the issue of complex application environments. It provides a flexible, agile way to unify data across processes and applications without coding. This means being able to integrate data and applications together whether they are in-cloud, on-premise or separated by geographical distance. It allows companies to access data and have it flow across the organization to users and systems as needed. Its capabilities include data integration, application integration, master data management, and reporting and analytics. It can be deployed as a point solution to integrate data between two applications or as a tool to unify all of a company's data and applications.

We have customers ranging from growth stage to Fortune 150.

Accur8 Software is one of CIO Review's Top 20 Most Promising Data Integration Providers for 2017.

For more information please visit our website at www accur8software.com or send an email to info@ accur8software.com

(1) Techopedia, <https://www.techopedia.com/definition/28008/application-management-am>