Accur8
SOFTWARE

Data Virtualization:
A Superior Data Integration Solution

BACKGROUND ON DATA INTEGRATION

Since the widespread adoption of computers in business over fifty years ago, there has been a need for companies to integrate data from different systems for reporting, analytics or application development purposes. In recent years, with the widespread adoption of cloud based and mobile applications, companies found themselves supporting an increasing amount of data from disparate data sources (silos) and formats. Keith Block, Salesforce.com COO, recently mentioned that "90% of the world data was created in the last 12 months". Faced with demands from management, users, and regulatory authorities to report or access this disparate data in a unified way, IT departments often took the expeditious approach of writing code or use connectors to integrate the required disparate data sources and respond to these requests. As a particular report or application proved popular, requests for additional data to be added often followed. This meant additional, time consuming, hand coding resulting in excessive drains on IT resources. Overwhelmed by these numerous demands, IT departments often found it necessary to delay responding, creating a backlog that resulted in an ever-more frustrated stakeholder base.

One solution that emerged in the early 1990s was the data warehouse: a central repository of data built from the disparate data sources of the business. This represented a tightly coupled architecture as all of the disparate data was physically stored and reconciled in one repository. With a single source of integrated data in place it was much easier for IT to satisfy a company's reporting needs. As significant of a move forward as this was, the shortcomings of the data warehouse approach become apparent:

- The creation of a data warehouse required incurring significant expense to setup what is a redundant, huge, physical relational database. A major component of this process was extracting, transforming and loading ("ETL") the data from the heterogeneous data sources into the data warehouse.

- Significant IT time and resource were needed to care and feed a data warehouse. If the underlying data sources were frequently updated, then the ETL process needed to be regularly re-executed to keep the data in the data warehouse synchronized.

- Not all of a company's data was stored in the data warehouse. If a report demanded a new data source then IT was required to devote time to integrate it into the data warehouse.

- Data contained in a data warehouse was not available in real-time which impacted the speed and quality of the decision leveraging that data. The currency of data extracted from a warehouse was only as current as the last time the warehouse's data were refreshed.

Data virtualization emerged as a technology to address the shortcomings of hand coding and data warehouse technology for accessing data from disparate data sources in an integrated way.

WHAT IS DATA VIRTUALIZATION?

Data Virtualization technology is rapidly gaining momentum and is radically improving the productivity of users and developers to access their company's distributed data sources for data integration, reporting and analytics, and application development. It is an agile method that enables real-time access to disparate data sources across the enterprise (on-premise and in the cloud) in a fraction of the time and at a fraction of the cost of traditional approaches. For those not familiar with the term, the following is a standard definition for data virtualization:

"Data virtualization extracts data from multiple disparate sources creating a unified virtual data layer that provides users easy access to the underlying source data."

Figure 1

**Graphic depiction of how data virtualization works.**
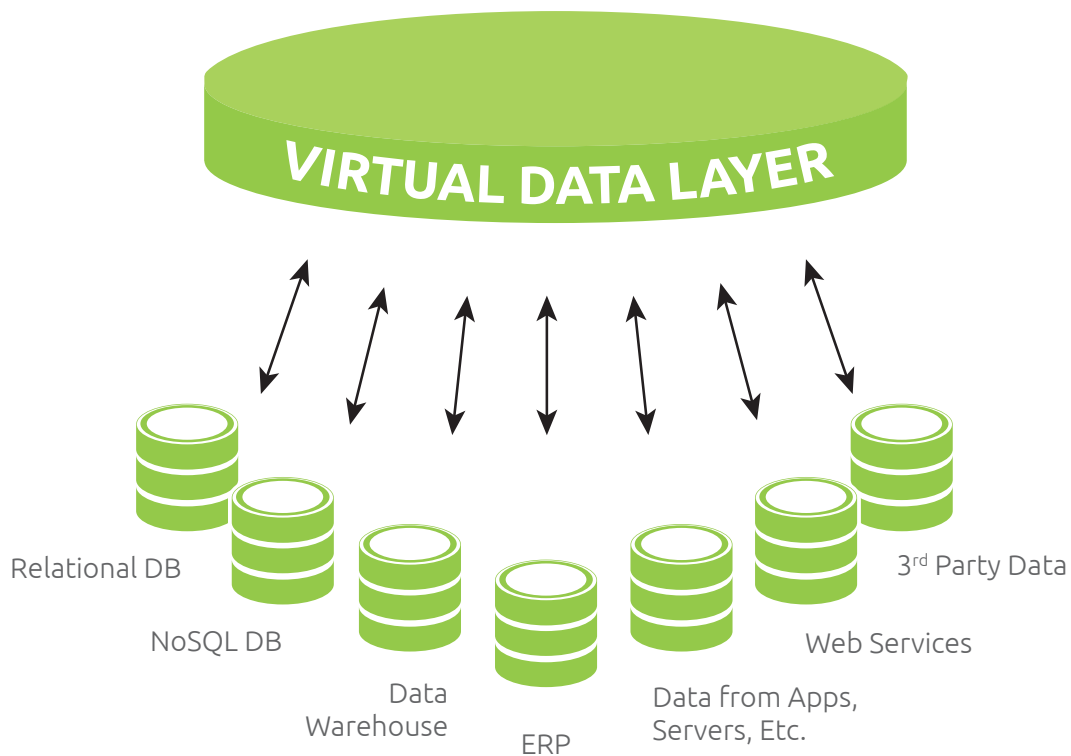
To access the data, a user (which can be a person or computer program) queries the virtual database, and it handles retrieving the required data from the underlying data sources. Importantly, this loosely coupled architecture means there is **no need to copy or replicate data from each constituent data source** to one repository as with data warehousing. Data virtualization provides the ability to transform the underlying data sources into unified forms that users can consume. It also offers the ability to create, update or delete information in the underlying data sources in real time.

03

Figure 2 below illustrates Accur8's technological approach to data virtualization. The first step is to model the data sources selected to be part of the virtual data layer to a metadata repository. This means the location of tables, the type of joins, security and other key bits of information are mapped from each data source to a metadata repository (essentially a warehouse that stores important knowledge about the underlying data sources. The resultant metadata repository effectively represents the DNA of a company's data and is a detailed map on how to locate all of the data that underlies the virtual data layer.

From here, Accur8's query engine comes into play. When a query for specific data is made to the virtual data layer, it is passed to our query engine that augments it with key information pulled from the metadata repository. This creates a transformed query that has the imbedded intelligence (location, security, business logic, auditing, etc.) to retrieve the required data from the underlying data sources. The query engine has the ability to seamlessly handle queries that require gathering data from two or more disparate data sources by stitching them together, in real time, into a single result set making it appear to the end user as if they are dealing with one virtual database. The end result is that, from the user's perspective, the virtual data layer provides seamless access to the underlying data as if it were one virtual database without needing to deal with the complexity of integrating each underlying data source. In fact, a developer doesn't even need to know where the underlying data is stored or located.

Figure 2

Detailed view of the mechanics of data virtualization
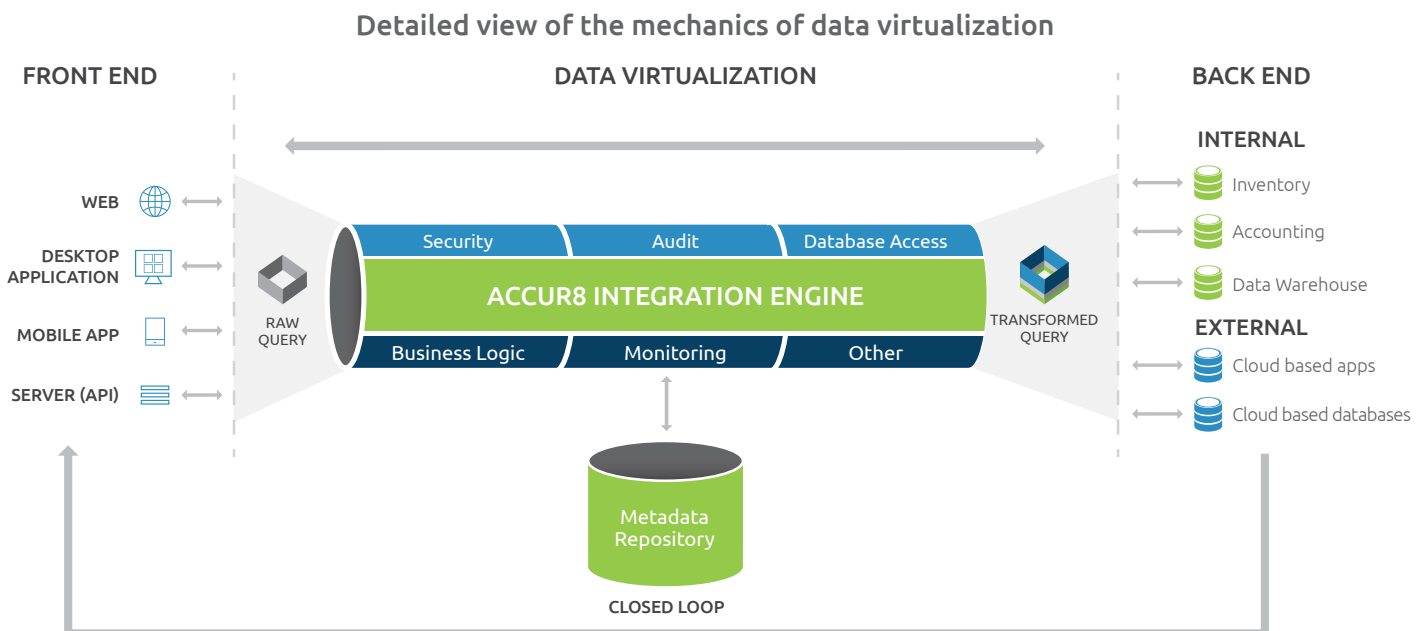
DATA VIRTUALIZATION VERSUS TRADITIONAL APPROACHES TO DATA INTEGRATION

An example of data integration would be a company wanting to improve the value derived from leveraging its CRM application used by its customer service representatives to manage and engage with customers. It wants to add important data from its marketing automation application and new

04

social media marketing application, so representatives can have a more holistic view of a customer when speaking with them. The discussions that follows explores the differences between the traditional approach of writing the code and connectors necessary to achieve integration and using data vitalization.

The traditional approach would be hand coding. This would require the following of a developer:

- To be familiar with the underlying data model of each data source.

- To be skilled in the development environment of each application (ie Ruby on Rails, .Net, Java, etc.)

- To write code that integrates each new application to the existing application. This would include the data modeling to determine the optimal way the various data sources will integrate. It would also require taking into account considerations such as the formatting, security, business logic, etc to properly integrate each data source.

- To test and deploy the developed code and ensure it is error free and meets requirements

The hand coding approach clearly works and would represent a hardwired relationship between the applications. However, there are certain shortcomings:

1. As described above, to properly integrate these two applications by hand coding requires significant amounts of time and skill from a developer.

2. If a new data source needs to be added or removed in the future it will require re-coding and retesting and will consume additional developer time to implement. It is this required investment in time that leads to initiatives getting delayed and IT backlogs. It is no surprising that typically, 80% of IT time and spending goes towards "keeping the lights on" vs working on improving the business.

3. The hand coded integration is brittle meaning it will likely break when there are future changes/upgrades to any software it is connected to. Increased downtime becomes a frustration for users.

4. To handle security it would require unifying the security of each application with an ad hoc security model which a developer may or may not have the skill and time to make robust. Hand coding creates many pathways between the existing application and the applications being integrated. More pathways equal more chance for security errors.

5. If there is a bug and the application breaks all that will be produced is an execution error. There will be no identification of the kind of error or its location. A developer will need to work through many layers (system logs, raw data, etc.) across different applications to figure out the problem before being able to cure it. If the developer that wrote the code has not provided proper documentation or has left the company this becomes even more problematic.

Figure 3

Difference between integration by hand code and data virtualization.

HAND CODING DATA   DATA VIRTUALIZATION
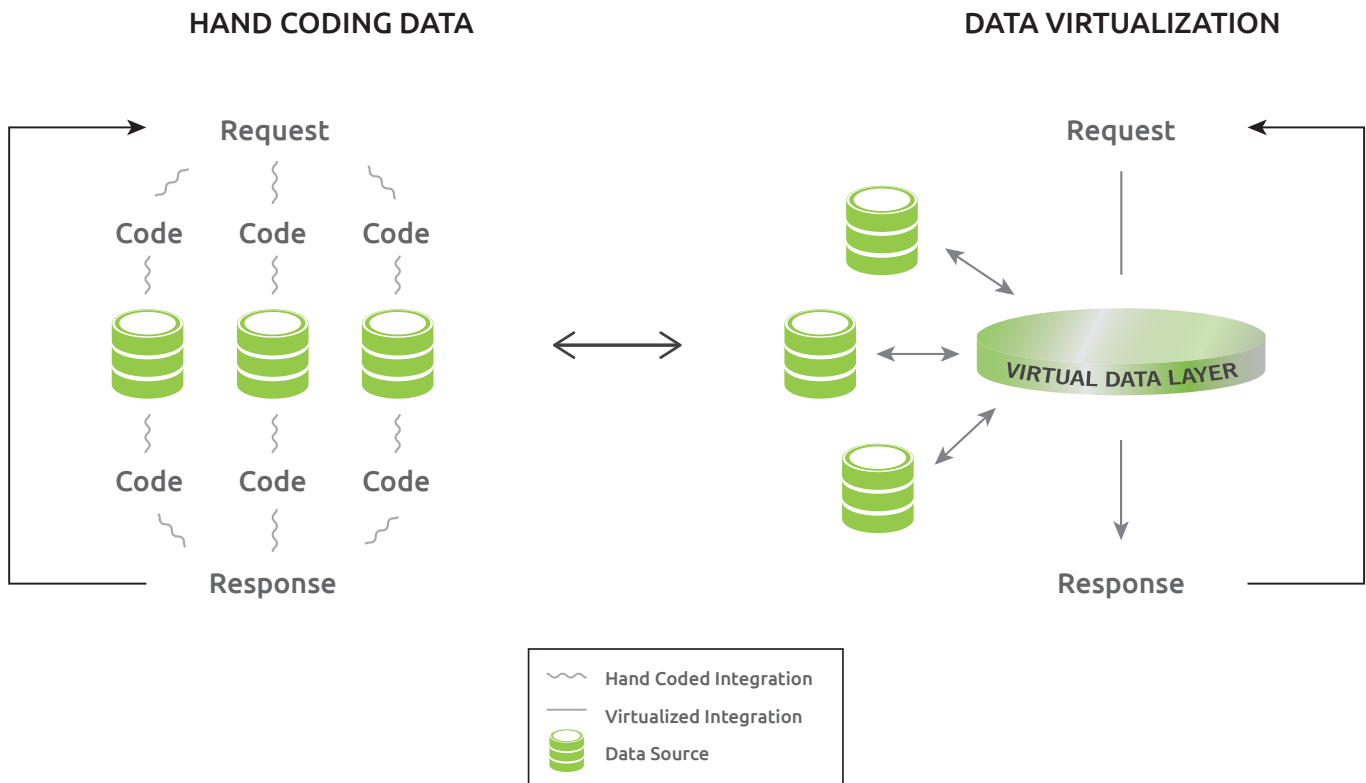
Alternatively, with data virtualization the databases underlying both the marketing automation and social media applications would be modelled and included in the virtual data layer. An initial virtual data layer could be completed in days. The final virtual data layer requires working with management to determine how they want their data modeled. The existing customer service application would then be integrated with the virtual data layer enabling it to retrieve data from each of the applications. This has multiple benefits including:

✔ **Easy Integration:** There is no need to deal with the complexities and workload of hand coding as described above. If the application requirements change and new data sources are needed they can be quickly accessed by the virtual data layer without re-writing code.

✔ **Improved Stability:** A virtual data layer can adapt to changes in underlying software that naturally happens over time due to things such as new versions being released.

✔ **Better Identification of Errors & Changes:** A virtual data layer provides specific notification of errors or any changes to its underlying data sources. This includes the description of the nature and location of the change. This allows a programmer to quickly identify and resolve an issue getting an application back up and running faster

06

✔ **Superior Security:** The virtual data layer provides one single access point to all underlying data sources and it retrieves data directly from a data source without needing to touch multiple layers of middleware as with hand coding. Users access can be tightly controlled in a fine grain or coarse grain way. Authentication, authorization and audit capabilities are all included.

✔ **Standardized Approach:** Data virtualization requires a structured process is followed to integrate data. This means consistent code and the ability to debug effectively. It also takes the variability of each developer's programming style out of the equation.

✔ **System Validation:** Should schema changes in the underlying data sources occur, schema changes can be proactively validated against the virtual data layer. The validation component reports on any feature no longer working.

✔ **Significant Savings:** For all the reasons mentioned above substantial money and developer time can be saved using data virtualization. The savings accrue not only in the initial set up but in ongoing maintenance. Some of our customers have experienced saving estimated at greater than 80% compared to traditional approaches.

EXPONENTIAL BENEFITS WHEN INTEGRATING MULTIPLE DATA SOURCES

There is an argument that for integrating a single, simple data source hand coding is the best approach. With a skilled programmer it can be done quickly and cost effectively. However, as the number of data sources required to be integrated grows and their complexity increases there are clear cut advantages for using data virtualization as all of the above mentioned benefits grow exponentially. Importantly, it also provides a very clear and understandable architecture compared to hand coding with can quickly grow into a labyrinth of brittle, unwieldy code if not managed correctly.

CONCLUSION

In conclusion, as data sources grow and become more complex inside each business the old approaches to integrating and accessing data are hitting a performance wall. Data virtualization is able to abstract data from source systems allowing a virtual data layer comprised of all underlying data sources to be easily built. This greatly streamlines a company's ability to access and consume its data while saving substantial time and money. It is certainly an approach every IT department should strongly consider.

Accur8
SOFTWARE

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