WHITE PAPER

INTELLIGENT BUSINESS STRATEGIES

Data Virtualization – Flexible Technology for the Agile Enterprise

By Mike Ferguson
Intelligent Business Strategies
February 2014

Prepared

Table of Contents

WHAT IS DATA VIRTUALIZATION?

In almost every business, data is becoming more distributed making it harder to access

In many organisations today, more and more business users are spending increasing amounts of time trying to find and integrate the relevant pieces of data they need to perform specific process tasks and to produce business insights. A major reason for this is because the data they need is increasingly becoming distributed across many different application data stores built to support specific functional business areas. This makes it difficult to access integrated, relevant data for any kind of cross-functional use.

When data is heavily distributed, people increasing start to integrate data using front-end tools and office applications

The impact of heavily distributed data is evident everywhere in organisations that have this problem. People struggle to access data and therefore contact others who know how to access the systems where the required data resides. The same thing happens when business users need access to data in applications they have no direct access to - perhaps because it is stored in applications specific to other departments. The result is that users 'collect' pieces of data that are passed back to them for manual data integration. These data fragments are often emailed as Microsoft Excel spreadsheets from which business users try to stitch data together using simple cut and paste to create yet more in Excel spreadsheets. This creates a common problem: when data is heavily distributed, business users often default to using front-end tools and office applications to integrate data when these tools and applications were never designed to perform this task.

Data virtualization software gives the impression that data is integrated when in fact it is not

In order to deal with this problem, data virtualization technology has emerged to simplify access to data. Data virtualization software gives the impression that data is integrated and stored in a database when in fact it is not. It just looks that way. Using this software it is possible to create multiple virtual views of data from multiple underlying internal and external data sources to present data as if it was integrated. When queried by applications and tools, this software then integrates the necessary data 'on-the-fly' at run-time to serve up integrated data on-demand. This is shown in Figure 1 below.

Data virtualization software integrates data at run time and serves it up on-demand

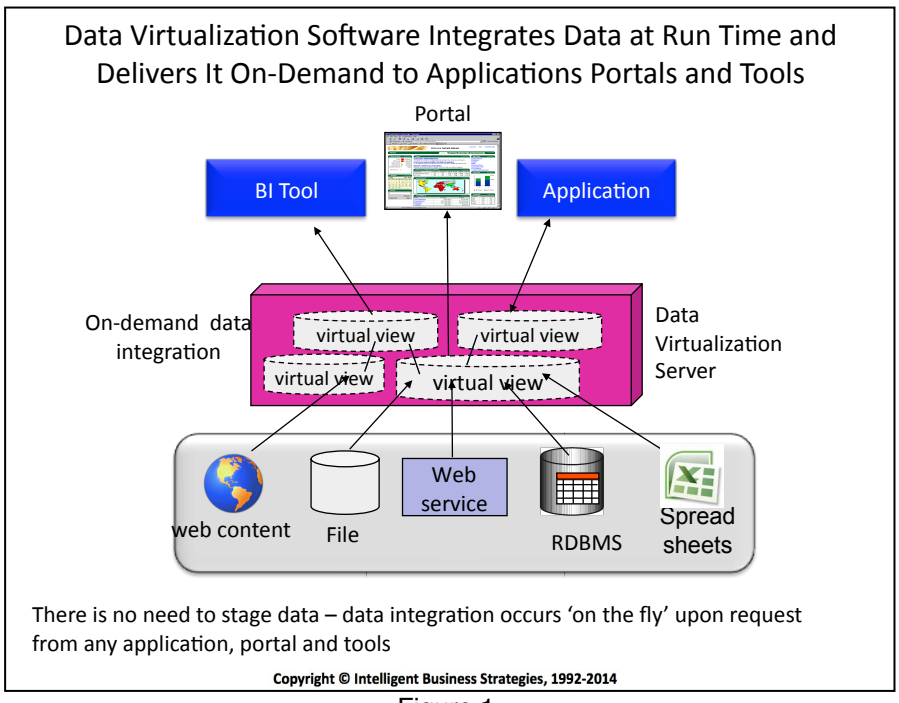

Figure 1

This paper looks at the increasing need for data virtualization software, how it works and presents popular ways to make use of it to maximise business value.

WHY DATA VIRTUALIZATION? – WHAT ARE THE BUSINESS DRIVERS?

Business users now want access to data that is stored inside and outside the enterprise

There are many reasons why data virtualization software is now becoming popular. The first, as already mentioned, is that data is becoming increasingly more distributed. Within many organisations, the distribution of data continues to grow unabated beyond enterprise and line of business applications already deployed in internal value chains. Today businesses are also adopting the use of multiple analytical appliances for specific types of analysis, Big Data stores, software-as-a-service (SaaS) applications on public clouds, social networking platforms, and accessing web and external data feeds to enrich on-premises data. All of this is making data more complex to access and manage as well creating more islands of data from where data needs to be integrated.

There are many reasons why disparate data in multiple sources may need to be integrated. These include:

There are many reasons why data needs to be integrated

- To provide data for query and reporting
- To simplify access to data in multiple operational systems
- To integrate core systems data with personal data e.g. Excel data
- To simplify access to multiple BI systems across the business
- To integrate on-premises and cloud data
- To create consistent data services across the enterprise
- To introduce agility into a data warehouse/BI environment
- To integrate CPM tools with multiple underlying DW/BI systems
- To provision integrated data into a portal
- To provide on-demand integration of distributed master data
- For heterogeneous replication

Data virtualization resolves the complexity of accessing highly distributed data

Just looking at the above list shows a range of business needs that are fuelling the insatiable demand for integrated data. All of the items in the list can be catered for using data virtualisation technology, which is why many organisations are turning to this technology to help them. Data virtualization technology offers flexibility while helping to resolve the complexity of accessing highly distributed data.

A classic use case for integrated data is the need to see across their value chain. Most organisations have a value chain. In manufacturing it is made up of two processes: procure to pay and order to cash. In other words, organizations need a view from supplier of materials, though materials inventory management, manufacturing operations, finished goods and product distribution. The manufacturing value chain is the entire operation. Similarly for logistics. In this case the value chain is from point-of-collection to point-of-delivery – the entire operation. In the oil and gas industry, it is explore-to-develop, build-to-operate, operate-to-market and market-to-distribute. No matter what business you are in, as you talk with middle and executive management, the need to see across multiple functions in a business operational value chain goes up as people need access to

information to manage the business at operational, tactical and strategic levels. Data virtualization provides that capability as shown in Figure 2.

Data virtualization helps organisations to manage their core busines operations by making it easier to see across a value chain

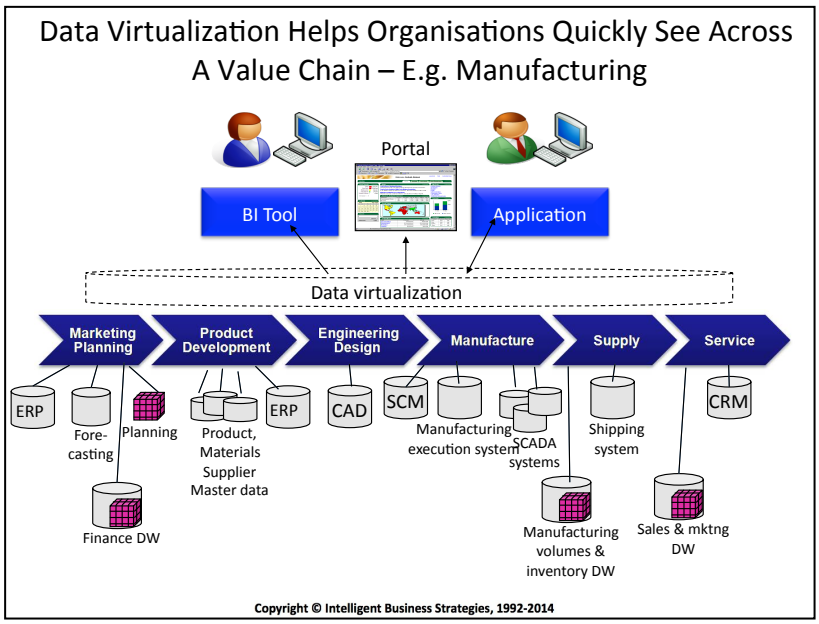

Figure 2

How Do Data Virtualization Products Work?

Data virtualization creates integrated virtual views of data in multiple underlying data stores

Data virtualization involves logically abstracting data from different data sources and data types into an integrated virtual layer consisting of multiple virtual data views. Data virtualization works by first defining common definitions for data and then using these to create integrated virtual views of data as if it was stored in a database. Virtual data views in this layer can be shared and queried on-demand to make consistent data available to all that need it when they need it.

Once this is done, a connection is made to disparate data sources by either accessing the data directly or connecting via an application interface. A wide range of data sources can be accessed including relational DBMSs, flat files, Excel spreadsheets, XML, web services, popular application packages (e.g. Salesforce.com), Big Data systems (e.g. Hadoop via Hive) and more.

Data virtualization servers connect to underlying data sources and map selected source data into integrated virtual views to make it look like data is stored in a single database

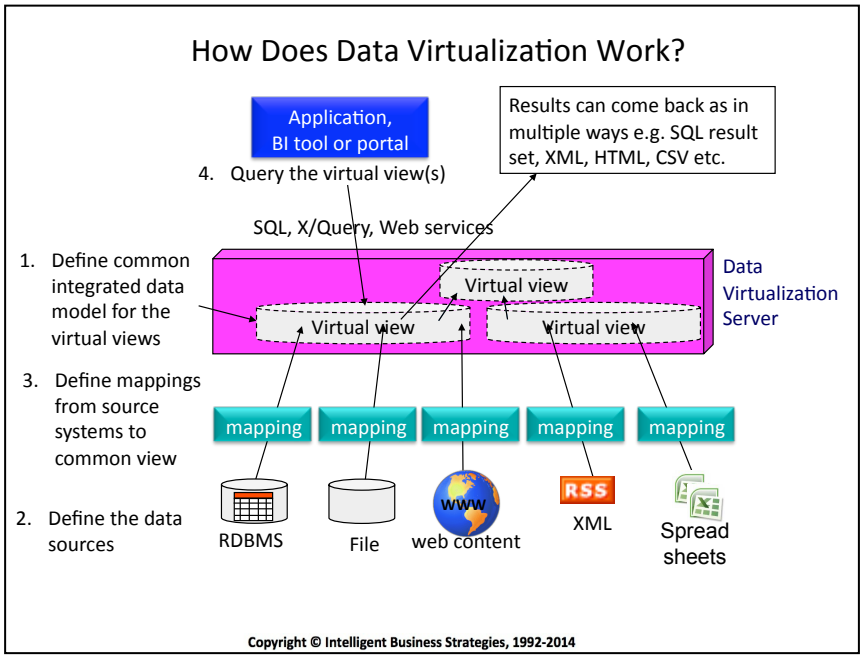

Figure 3

Data can then be queried as if it was in one database

Data of interest in source systems can be mapped to the virtual views in the data virtualization server to tell the data virtualization server how to integrate the data needed to provision the virtual views. Once all mappings are defined, they are stored as metadata in data virtualization server metadata repository and the virtual data structures are available for use by applications and BI tools. The three step process of setting up a data virtualization server is shown in Figure 3.

The data requested is then integrated on-demand at run time

At this point applications, portals and BI tools can query the virtual views as if the data was integrated using SQL or in some cases other query languages like X/Query. Querying the virtual data structures results in on-demand data integration without having to replicate or physically move data residing in multiple underlying data stores. Instead, a data virtualization server will federate the query such that parts of the query are sent to the appropriate underlying data stores and/or applications to retrieve the necessary data that the requesting application or tool needs. This appropriate data is then retrieved and integrated 'on-the-fly' before sending back the results to the requestor in the appropriate format.

POPULAR DATA VIRTUALIZATION PATTERNS

There are many data virtualization patterns which makes this technology very versatile

Given the flexibility of data virtualization technology, there are many usage patterns that exist. Some of the most popular are data virtualization patterns for use in:

- Business Intelligence (BI) and Corporate Performance Management (CPM)
- Data warehousing (DW)
- Master Data Management (MDM)
- The creation of information services
- Cloud Computing
- Data management, including data quality and data integration

The use of data virtualization in support of **BI and CPM** includes:

- The ability to integrate CPM scorecards and managerial dashboards with multiple underlying functional BI systems (e.g. Marketing DW, Finance DW, Supply Chain DW, Operations DW) to calculate cross-functional KPIs. This very common requirement is shown in Figure 4.

Calculate KPIs from data in multiple underlying data warehouses

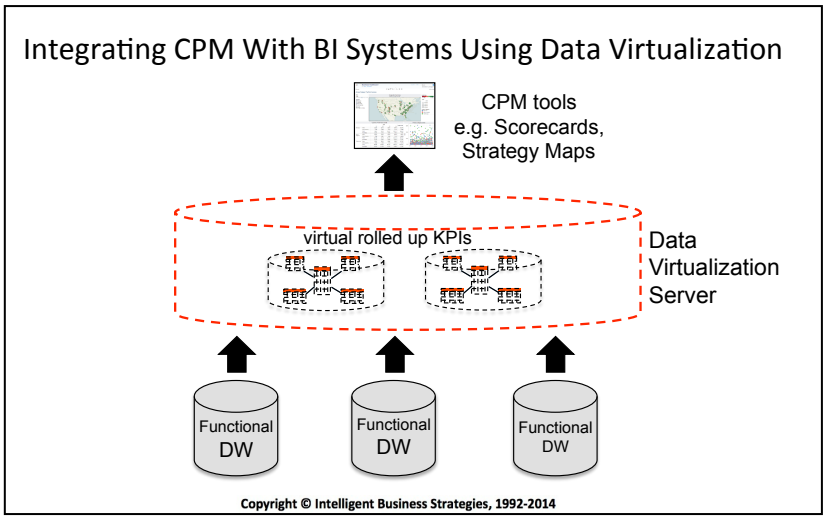

Figure 4

Integrate live and historical data on-the-fly

- The ability to integrate data from multiple underlying operational systems into a virtual ODS for reporting
- The integration of up-to-date live operational data in multiple transaction processing systems with historical data in a data warehouse for 360° holistic reporting e.g. for a complete picture of customer activity

Simplify and rapidly implement regulatory reporting

- The ability to integrate data from multiple underlying operational systems for regulatory and up-to-the-minute operational reporting

With respect to **data warehousing,** the construction of data warehouse databases, two major uses of data virtualization stand out. The first is the use of data virtualization to create virtual data sources to simplify ETL access to data residing in multiple data sources. By using data virtualization to create virtual data sources, ETL tools can extract data from virtual views, thereby shielding ETL jobs from structural changes in underlying data sources. The second data warehouse pattern is the use of data virtualization to create virtual data marts as shown in Figure 4.

Virtual data marts simplify data warehouse architectures making it faster to deliver value and reducing the total cost of ownership

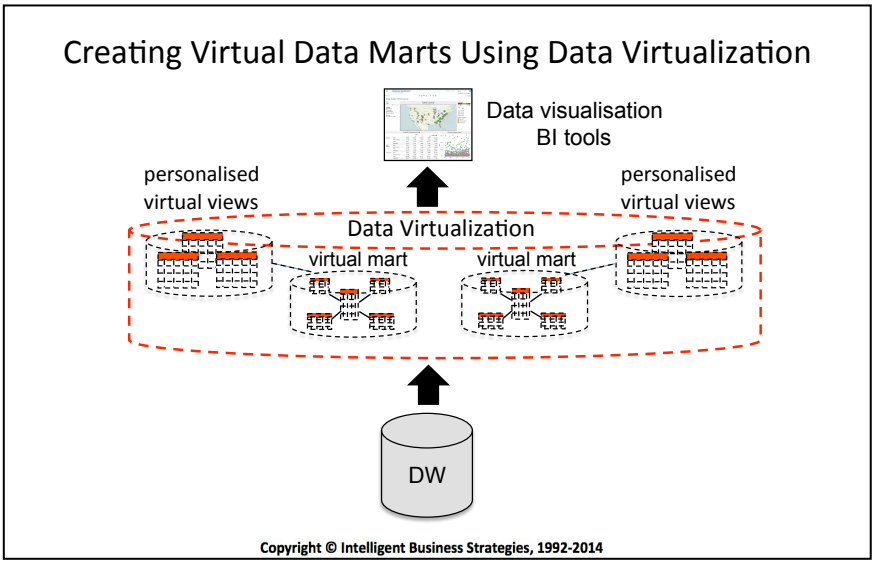

Figure 4

Data virtualization can also be used to dynamically integrate **master data** for consumption by operational applications and data warehouses. This is referred to as virtual master data management (MDM) and is also known as the registry approach.

Introduce information 'services'

One of the most powerful capabilities of data virtualization is the ability to store popular federated queries on data virtualization servers and publish them as web services. These are often referred to as **information services**. These are discussed in more detail later in the paper.

Integrate cloud and on-premises data

With respect to **cloud computing**, popular cloud-based data sources like the Salesforce.com application and others can be treated like any other data source. In this case the data virtualization server simply connects to off-premises cloud-based data stores and applications to integrate cloud and on-premises data. This is shown in Figure 5. For example, sales leads data from salesforce.com could be integrated with Google geocode data to see where the leads are located in a mapping application.

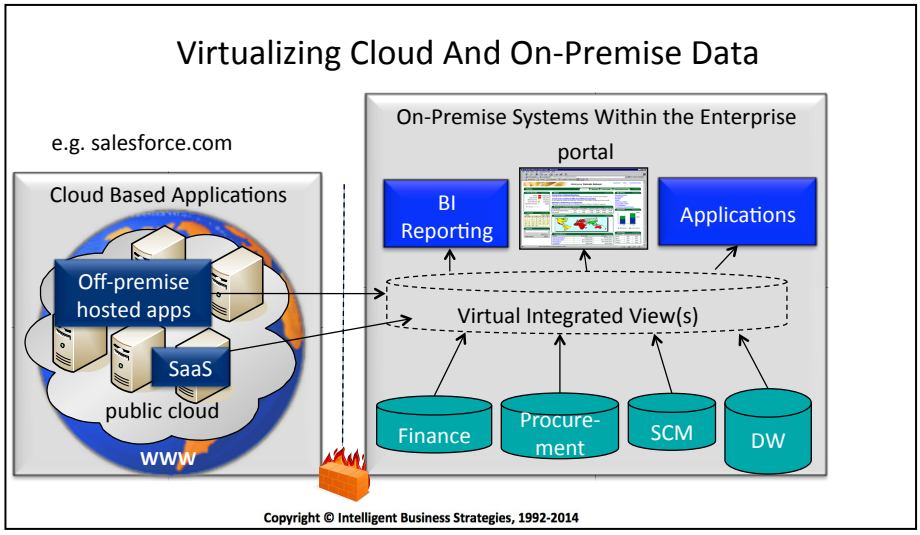

Figure 5

THE IMPACT OF DATA VIRTUALIZATION ON THE ENTERPRISE

Having looked at what data virtualization is, how it works and popular usage patterns, a natural next question is "What kind of impact can data virtualization have on an organisation?" This question can be answered by looking at how this technology:

- Impacts data warehouse architecture
- Introduces agility to simplify self-service BI
- Is able to cope in terms of performance
- Fuels the creation of shared departmental and enterprise data services
- Helps manage master data
- Helps or hinders enterprise data governance

THE IMPACT ON DATA WAREHOUSE ARCHITECTURE

The introduction of data virtualization into any organisation can have a significant impact of existing data warehouse architecture. For the last twenty years we have been building data warehouses in a 'waterfall' approach where data is integrated into an operational data store, then into a data warehouse. From there, detailed data is aggregated again using data integration tools, to create physical data marts. Finally we aggregate data again to populate OLAP cubes to support multi-dimensional analysis. The problem with this approach is the length of time to value. The focus is on building an ODS, then a data warehouse, then physical data marts, then cubes and finally business users get to access the data. It is not surprising with this approach that business users get frustrated waiting for the completion of all of these layers.

The traditional approach to data warehousing is focussed on building data warehouses first and offering business value second

Additional issues arise with this traditional method to building data warehouses including that of complexity. There are multiple physical data stores. An ODS, a data warehouse, multiple physical data marts and multiple physical cubes stored in an MDBMS. The total cost of ownership in managing this complexity can be significant. To give an example, physical data marts require hardware server(s), a DBMS software license, and developers to design a data model and build ETL jobs to populate these marts. The other major problem with this approach is the impact of change. New business requirements can result in a 'domino effect' when it comes to changes needed to accommodate these requirements. Changes may be need to data warehouse, data mart and cube data models. ETL jobs associated with all of these may also have to change. In addition, resource constrained IT professionals may become a bottleneck and so agility is limited.

Data virtualization reverses this by focusing on delivering business value first

The introduction of data virtualization into data warehouse architecture has a number of benefits. This approach speeds up time to value and simplifies architecture because fewer physical data stores need to be built. This is because virtual data marts can be introduced instead of physical data marts thereby simplifying the overall architecture. Figure 6 shows how data virtualization fits into data warehouse architecture. Note, this is not a replacement for a data warehouse.

Data virtualization means that fewer physical data stores need to be built

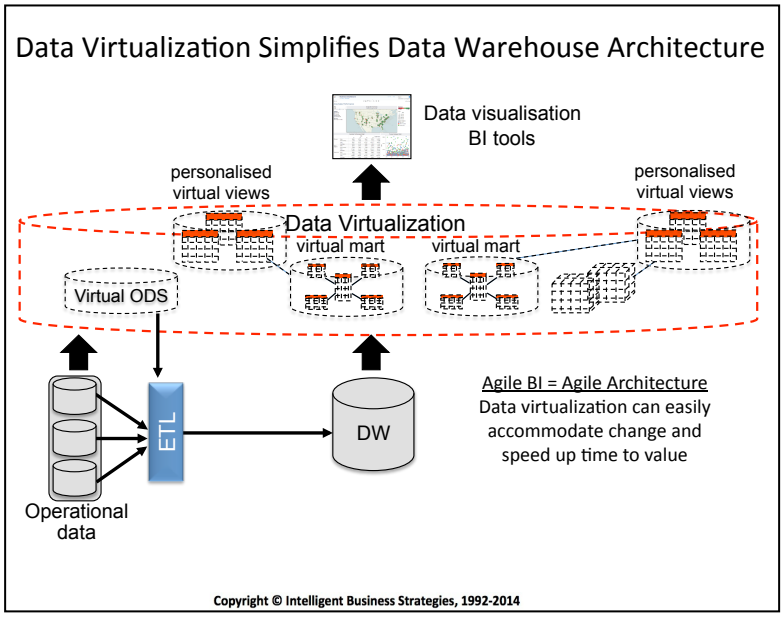

Figure 6

Data virtualization simplifies data warehouse architecture

Introducing data virtualization into data warehouse architecture allows organizations to create a virtual ODS, virtual data marts and virtual ROLAP cubes thereby reducing the number of data stores needed. It also speeds up time to value because fewer data stores need to be built which in turn reduces the cost of ownership involved in building and maintaining data warehouse architectures.

Data virtualization facilitates early prototyping

Business users are engaged at an early stage which increases the likelihood of success

In addition to the architecture benefits, the creation of virtual data marts facilitates early prototyping even if data is not yet in a data warehouse. This is because it becomes possible to create virtual views that integrate operational data and historical data as data virtualization servers can connect to operational transaction processing systems (or replica copies of them) and data warehouses as data sources. Engagement of business users at an early stage is always beneficial in a data warehouse project because BI can be produced and worked on with business users even if some data has not yet been brought into the data warehouse. Once users are happy with the prototype it also means that this can be released to provide business value while effort then switches to bringing any necessary data into a data warehouse to extend it. This 'prototype first, build second' approach is much more focused on time to value as opposed traditional approaches that focus on IT build first and offer business user access last.

DATA VIRTUALIZATION – ADDING AGILITY TO EASE SELF-SERVICE BI

Data virtualization can easily accomodates change making it an agile technology

Rapid prototyping is one way of increasing agility during data warehouse development. However there are a number of other ways to do this, all of which help facilitate self-service BI. The first example of this is accommodating change. Figure 7 shows an example of how a new column can be added to a virtual table with ease. This allows virtual structures to be changed and tested quickly to accommodate new business requirements thereby speeding up time to value. Contrast this with traditional approaches that may cause changes to multiples data store data models and changes to ETL before a business user can see the benefit. In addition, data virtualization allows new data sources and new virtual tables to be added quickly and easily.

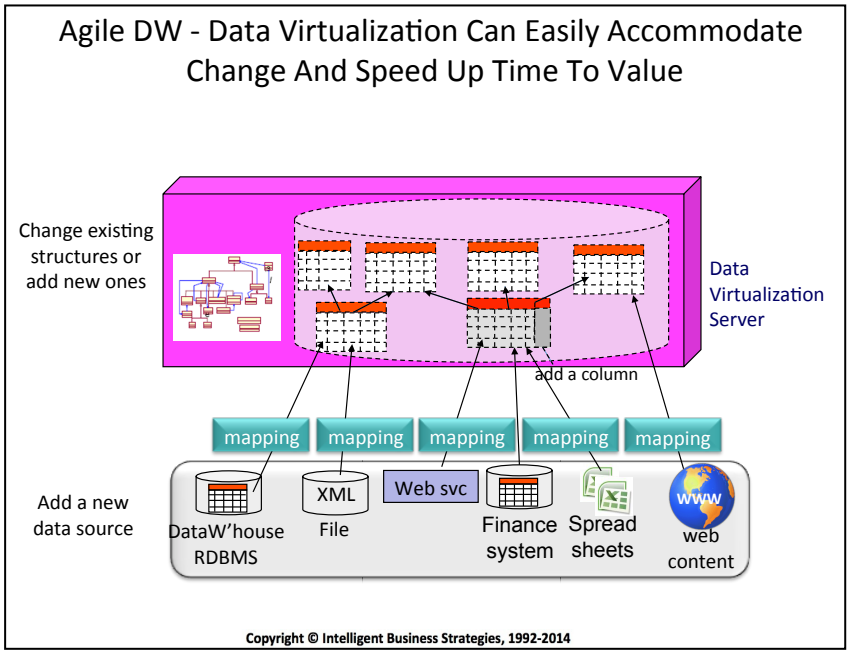

Figure 7

Data virtualization increases the likelihood of success with self-service BI

How does this facilitate self-service BI? A critical success factor of self-service BI is the ease of access to information. Almost always, business users want access to multiple data stores whether it is multiple data warehouses, data warehouse and transaction processing system data or data mart and spreadsheet data. A common feature of easy-to-use data discovery and visualization self-service BI tools is that they can access multiple data sources. However, the problem that this causes is that information producers on the business side start to use front-end BI tools to do their own data integration. In other words, self-service BI turns into self-service DI (data integration) using tools that were never designed to do data integration.

The second example of adding agility is the use of data virtualization to hide the complexity of accessing multiple underlying data stores from end user information producers in order to simplify their access to data. It makes data easier to access by integrating it on-demand, removes the need for self-service DI and allows business users to focus their time on analysing data and producing business insights rather than getting caught up in the complexity of multi-source data integration. It also means that integrated data is consistently made available to all self-service BI users thereby avoiding an anarchy situation whereby many users have inadvertently calculated the same metrics differently.

MANAGING PERFORMANCE WHILE ACCOMMODATING FLEXIBILITY

In order to maintain performance with data virtualization, frequently accessed virtual data structures can be cached on the data virtualization server so that repeated retrieval from frequently access data in underlying data sources is avoided. This is shown in Figure 8. Also if data needs to be integrated and frozen at a particular point in time, it is possible to create 'snapshot' virtual tables whereby data integration occurs at a particular time (e.g., once a day, once per hour etc.,) and cached for subsequent data access without the risk of the data changing. The point here is that *both* dynamic *and* snapshot data integration are possible. It is not one or the other.

Data can be cached in memory to improve performance

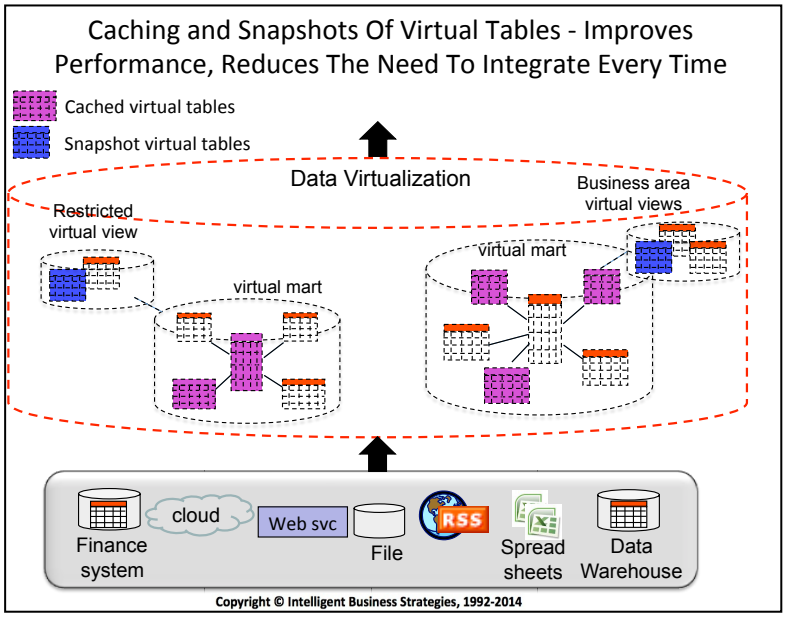

Figure 8

DATA VIRTUALIZATION – INTRODUCING DATA SERVICES IN A SOA

Popular federated quers can be stored on the server and invoked on-demand

In order to guarantee consistency of data, popular federated queries can be saved on the data virtualization server and published as web services to serve up data on-demand (in various formats) to consuming applications. In some cases, virtual data services can also be updated to support transaction processing across multiple underlying systems.

With this capability organisations can standardise departmental and enterprise information services to consistently deliver the same information to multiple applications and tools. An example of this is shown in Figure 9 with four information services that can be made available on an enterprise service bus (ESB) and invoked on demand by other applications and portal to provide the required data on in-progress orders, finished goods, shipments and payments due.

Publishing federated queries as information services means the same information can be consistently delivered to multiple applications and tools

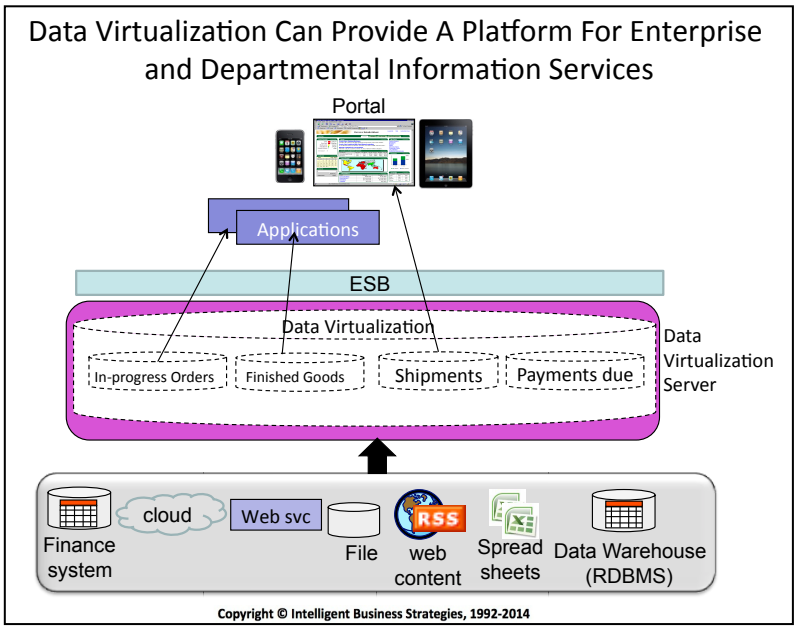

Figure 9

The benefit of this is that it encourages re-use rather that re-invention so that applications and self-service BI users are consistently supplied with the same data.

DATA VIRTUALIZATION AND MASTER DATA MANAGEMENT

Data virtualization can also play a role in master data management in two ways. As long as it is possible to support data cleansing, define data matching rules and support global identifiers, data virtualization software can be used to implement virtual MDM. This is shown in Figure 10 and is sometimes referred to as the registry approach

Virtual master data management is a low cost way to get started with MDM

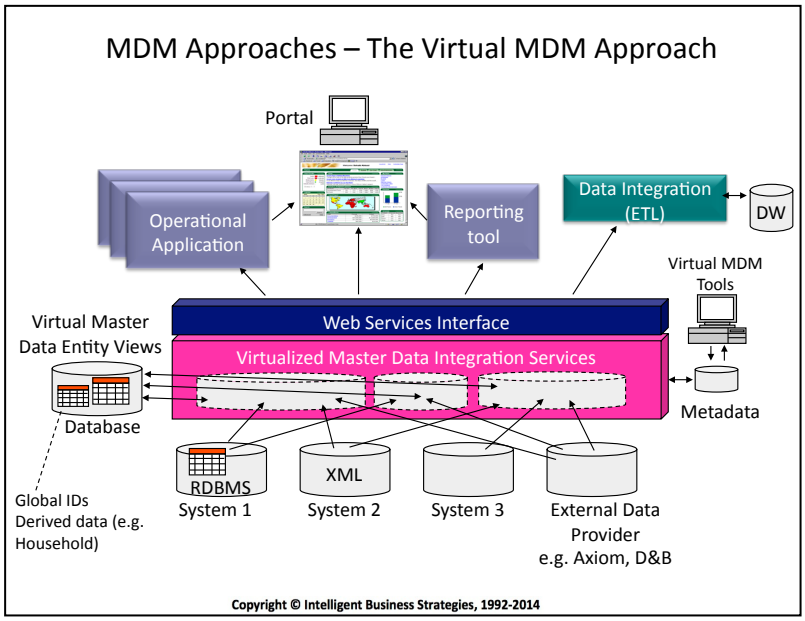

Figure 10

Data virtualization can also add value to existing centralised MDM solutions

For organisations that have consolidated master data in a centralised data store as a system of record (SOR), data virtualization may still have a role. This is because there are often shared master data attributes (e.g. customer, product) that are consolidated while master data attributes specific to only one application may still remain in that system. In this case data virtualization can provide a holistic view of a master data entity by integrating data from a centralised MDM hub and any other applications still holding master data attributes that are not shared.

THE IMPACT OF DATA VIRTUALIZATION ON ENTERPRISE DATA GOVERNANCE

Data virtualization can also be used to help implement data governance

With respect to enterprise data governance, data virtualization can make a significant contribution in helping to govern data especially when a data virtualization server is part of an Enterprise Data Management platform (tool suite) sharing the same metadata repository as other data management tools in the suite. If data virtualization is part of a data management platform that also includes a business glossary, a data modelling tool[1], data discovery, data quality profiling and cleansing software, then best practices can be introduced as shown in Figure 11. In this case, business users can define common data names and definitions for master and transaction data in an enterprise business glossary to create a shared business vocabulary (SBV) for the enterprise. These common data definitions can then be used to create a logical data model that can be used to create virtual tables in a data virtualization server. The result is that all applications and tools accessing virtual views in a data virtualization server will see data described using common data definitions. In addition data quality can also be enforced.

[1] It may be the case that the data modelling tool comes as part of the data virtualization software.

Data virtualization can share metadata with other tools in a data management platform

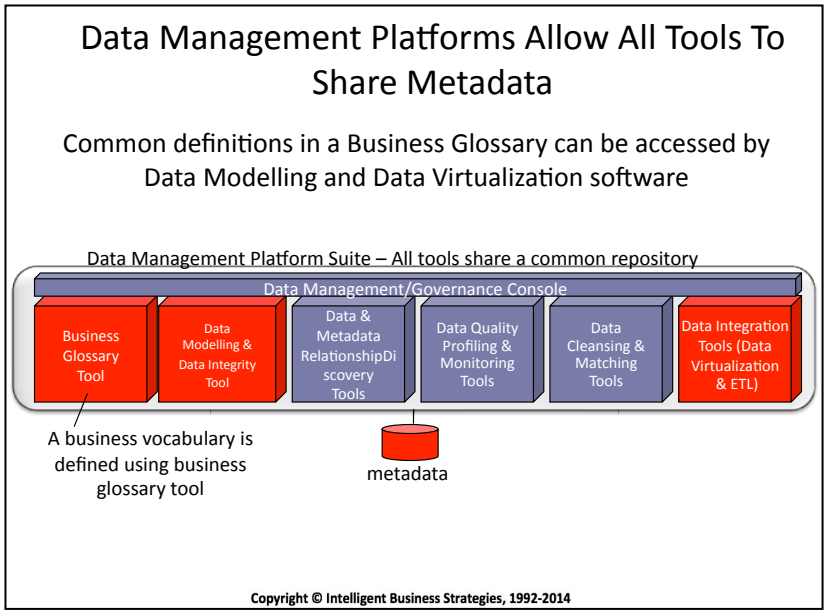

Figure 11

Common data names defined in a busness glossary can be used in virtual views on a data virtualization server

Pointing BI tools at a data virtualization server that uses common data names drives consistent common definitions across all BI tools, cubes, dashboards and reports

For organisations that have multiple BI tools it means that all tools with semantic layers will have consistent common SBV data definitions. Similarly self-service BI tools without any semantic layer will also see the same common data definitions when accessing virtual data structures that serve up integrated data.

Figure 12 shows three further ways in which data virtualization technology can help implement data governance while also simplifying business user access to data in self-service BI environments. The first of these is the creation of shared data services by storing federating queries on the data virtualization server and publishing them as web services. This has the effect of standardising data access. The second is by connecting the data virtualization server to an integrated master data management systems and using it as a data source. This simplifies data integration. Finally, the use of nested virtual views also helps govern data because views on views of data can be used to restrict access to data to only those people authorised to see that data. This capability also provides flexibility in that multiple layers of nesting can offer up many different views of integrated data to satisfy the needs of different users and applications while protecting data at the same time.

Common data names, common information services, resticted access to data and master data management all play their part in helping to govern data

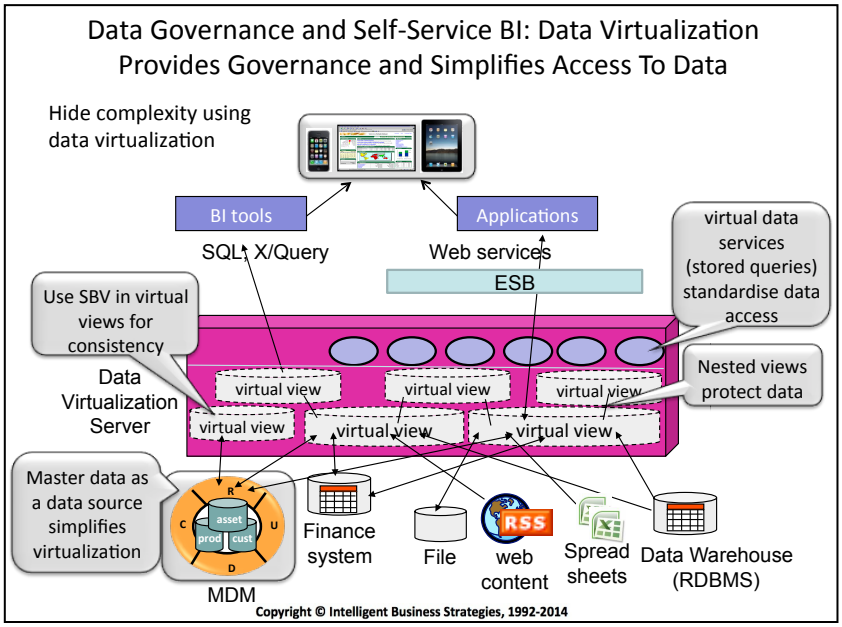

Figure 12

REQUIREMENTS FOR DATA VIRTUALIZATION TECHNOLOGY

Having seen the benefits and flexibility that data virtualization technology can offer, what kinds of requirements should you consider with respect to this software if you are looking to add this technology to your data management toolbox? The following is a set of requirements worth considering. They are not listed in any particular order but are categorised for ease of reading. It is up to the reader to define which of these are more important than others. Also note that the list is not intended to be exhaustive nor is it intended that all of these should be considered mandatory.

Data virtualization software should be able to connect to multiple data sources

- Data virtualization connectivity requirements
  - Connectivity to a wide range of data sources including:
    - Relational DBMSs e.g. via JDBC, ODBC or X/Query
    - Flat files
    - Multi-dimensional DBMSs (via MDX)
    - Non-relational DBMSs e.g. IBM IMS, Software AG Adabas
    - XML data sources e.g. XML files, RSS feeds, MS Office documents
    - Message queues
    - Web services
    - Content management systems
    - Documents, PDFs
    - Big data stores e.g. Hadoop Hive and/or HDFS
    - Popular packaged applications e.g. SAP, Oracle E-Business Suite, Salesforce.com
    - Cloud data sources
    - LDAP directories
    - Microsoft SharePoint lists
- Data virtualization base functionality requirements include support for:

Graphically define virtual data views and map source data to them

  - Graphical mapping of data from multiple disparate data sources to virtual data structures defined to the data virtualization server
  - Adding new virtual structures and changing existing virtual data structures on the fly to accommodate new business requirements
  - Adding new data sources on the fly to accommodate new business requirements

You should be able to make changes on the fly

  - Nesting virtual structures on top of one another to allow for flexible access to data and to restrict data access to integrated virtual views
  - Enforcing data quality policies when integrating data on-demand to provision high quality data
- Virtual data access requirements include data virtualization server support for:

- The SQL-92 API
- Either a SOAP or REST web service API to permit applications and tools to invoke federated queries (stored on the server) on-demand
- Other interfaces to open up access to a wider range of applications e.g. X/Query

SQL access and web services support are key requirements

- Query result requirements
  - Ability to define one or more formats in which query results could be returned e.g. SQL results, XML, CSV format, HTML format, JSON format, Excel format
- Performance requirements include the ability to support:
  - A cost-based optimizer
  - Access data management platform metadata or underlying data source metadata to know where the data is located in underlying systems and what type of data store the underlying data is stored in
  - Caching of popular virtual table structures in memory to speed up access to data
  - Scheduling of data caching on a timer-driven basis to govern when data should be materialized and also to define policies as to when (if at all) it should be refreshed
  - Snapshot data to freeze data at a particular point of time
  - Clustering a data virtualization server to scale it beyond a single instance of the server
  - Monitoring the performance of the data virtualization server itself and to make configuration changes to scale the server

Federated query optimization and data caching will help performance

- Security requirements include the ability to:
  - Integrate a data virtualization server with a corporate LDAP directory
  - Authenticate users, applications and service requests through a common security service
  - Restrict applications, tools and user access to virtual data structures and data services on a data virtualization server via a common authorization service

Secure access to data should be supported

- Data management platform integration requirements include the ability to:
  - Integrate with business glossary technology and/or data management tools with a data virtualization software to allow common data definitions defined in a business glossary to be used to define virtual data structures
  - Import data models (preferably with common data definitions) from data modeling tool(s) to use as a virtual data model in the data virtualization server
  - Reverse engineer data structures on source systems where appropriate and/or to use data discovery software to discover what data is held there

Integration with business glossaries, data modelling and data quality software is a key differentiator

DATA VIRTUALIZATION PRODUCT EXAMPLE: SAS FEDERATION SERVER

Having looked at the capability and flexibility of data virtualization and defined requirements that this type of technology should support, this section looks at how one vendor, SAS, has stepped up to meeting these requirements to help customers get the most out of this technology.

SAS Institute is a recognized leader in providing enterprise business analytics and data management. Headquartered in Cary, North Carolina, SAS has been providing business analytics and statistical analysis software since 1976.

SAS Federation Server is a new addition to the SAS data management suite of technologies.

SAS has 65000 customers worldwide, and its distribution channel spans 140 countries. SAS has strategic alliances and partnerships with leading hardware and software vendors, and regional and global solution delivery partners. SAS provides a data management platform part of which is the Federation Server data virtualization software.

SAS Federation Server offers applications, portals and BI tools the ability to query data from disparate underlying data sources with a single SQL query. This is done by creating of a layer of abstraction over physical data that may be stored in a variety of data sources. Data virtualization capabilities available in SAS Federation Server are listed below and are categorised in a similar fashion to the aforementioned requirements for easy reading. They include:

SAS Federation Server connects to popular databases

Data virtualization connectivity

- Centralised set-up and maintenance of data source connections
- Connection to a range of data sources via SAS Access native and ODBC drivers. This includes connectivity to Oracle, DB2, Teradata, SQL Server and MySQL database management systems
- Support for connection pooling between the SAS Federation Server and source data systems

SAS Federation Server also provides data quality support

Data virtualization base functionality

- A graphically query builder to create materialized integrated virtual views of data
- The ability to change virtual views on-the-fly
- Integration with SAS Data Management Platform to enforce data quality policies when integrating data on-demand

Applications can access virtual data structures via SQL

Virtual data access

- A SQL interface (via JDBC and ODBC) to virtual data structures that trigger on-demand data integration from multiple data sources
- Support for statement pooling from the SAS Federation Server JDBC driver

Query results

- Query results can be returned as SQL results

Performance

- Federated SQL optimization - SAS Federation Server can query metadata in underlying source databases to build an optimized query plan on how to best perform SQL operations

Federated query optimization and data caching are available to help improve performance

- Ability to enable or disable data caching of materialized views
- Scheduling of data cache refreshes
- Snapshot data cache
- Ability to monitor the current and historical status of all sessions connected to the Federation Server
- "Native" database drivers to underlying data sources (often faster than traditional ODBC access)
- Automatic threading of query processes, which allows transformation and integration work to be done in parallel

Security

- Ability to define access permissions that control user and application access to data
- Nested and multiple virtual views of data from underlying data sources to restrict access to sensitive data

INTEGRATION WITH SAS DATA MANAGEMENT

SAS Federation Server also integrates with other tools in the SAS Data Management suite of products including SAS MDM

In addition to the functionality within SAS Federation Server, it also has the ability to integrate with other components within the SAS Data Management suite of products.

However SAS Federation Server also integrates with SAS MDM to support master data management initiatives. Integration occurs in a number of ways. The first is that SAS Federation Server can be used to simplify access to data sources that hold fragments of master data by creating virtual master data sources and making these available to SAS MDM. This means that SAS MDM jobs can access virtual master data views to consolidate and integrate that data into a centralised master data store. In addition, scheduled caching of source data from SAS Federation Server can make the right data available to SAS MDM at the right time. Figure 13 shows two uses of SAS Federation Server with SAS MDM.

The first is in providing data sources and the second is in providing restricted views of master data to applications to hide sensitive data. It may also be the case that not all data for a master data entity is in a master data hub in which case a 360 degree virtual MDM view can still be created by making the SAS MDM master data hub and other source systems data sources to SAS Federation Server. In other words it is possible for SAS Federation Server to sit below and above SAS MDM.

It is also possible for just one of these options to be deployed to create a virtual MDM solution whereby data elements can be omitted from standard SAS MDM tables and views by defining the master data layout in SAS Federation Server. Data elements can be hidden by column, by row or both allowing precise control of sensitive master data made available to other enterprise systems.

SAS Federation Server can simplify access to distributed master data and integrate SAS MDM master data with other data sources

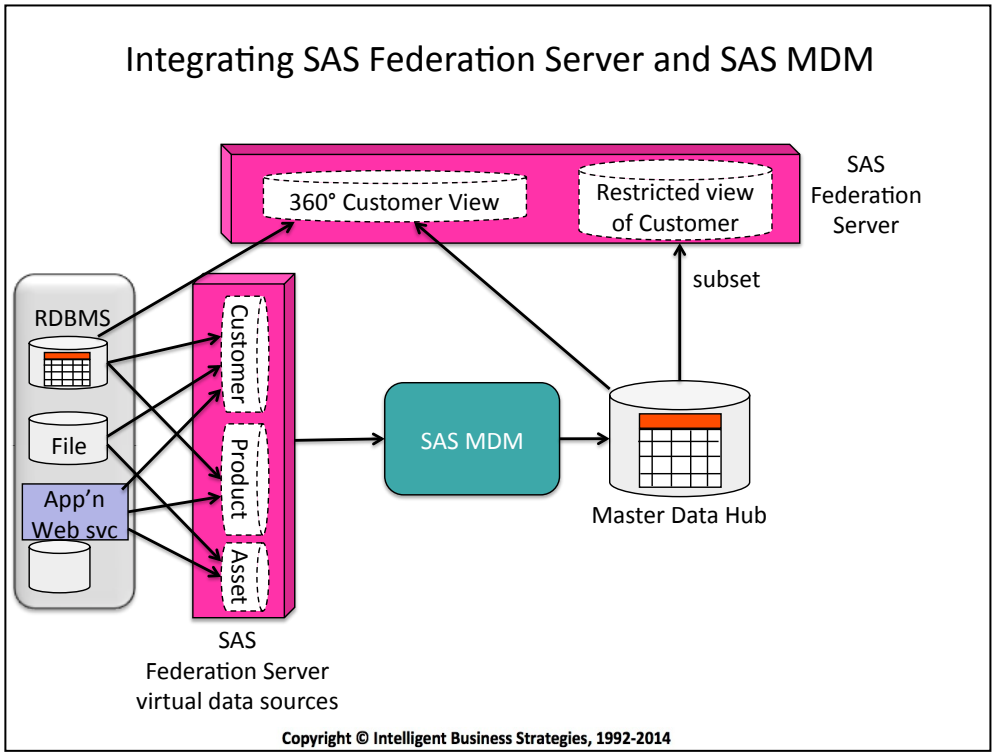

Figure 13

DATA VIRTUALIZATION AND SAS ENTERPRISE BI SERVER

SAS Federation Server simplifies access to data from SAS and other third party BI tools including

Given that SAS Federation Server supports any application, portal or tool wanting to access data in multiple underlying data stores, a powerful combination is the use of SAS® Enterprise BI Server with access to an underlying SAS Federation Server.

SAS Federation Server makes it easy for users of SAS Visual Analytics self-service BI tool to access data

SAS® Enterprise BI Server is a comprehensive, business intelligence software platform (suite of tools). It includes multiple role-based, BI tools aimed at different types of users within a well defined IT governance framework and a centralized point of administration. To cope with the increase in demand for self-service BI within many enterprises, SAS has added SAS Visual Analytics to its product portfolio of BI tools within the SAS Enterprise BI Server. SAS Visual Analytics is a high-performance, in-memory solution for exploring large amounts of data very quickly. It enables users to spot patterns, identify opportunities for further analysis and convey visual results via web reports or mobile devices. A key component of SAS Visual Analytics is the SAS® LASR™ Analytic Server which reads data into memory for SAS Visual Analytics users to access. This speeds up analytical processing.

However, a key success factor to self-service BI is easy access to data. By making integrated data from multiple underlying data stores available via SAS Federation Server, it becomes possible for SAS® LASR™ Analytic Server to easily access integrated data and bring it into memory for SAS Visual Analytics users to analyse. Pre-built data services (federated queries) stored on SAS Federation Server simplifies access to data even more while also guaranteeing that self-service BI users don't re-invent data integration when wanting access to the same data. The use of data virtualization to hide complexity of data access from BI users is a very powerful combination and something well worth making available.

CONCLUSION

The increasing trend in the need to access distributed data is driving up demand for data virtualization

The increasing trend of data distribution and the complexity it brings means that data virtualization is now an essential tool in a company's enterprise data management toolkit.

Data virtualization is quick to implement and easily accomodates change

Data virtualization software is quick to implement, easily accommodates change (agility) and can simultaneously deliver different data for many different uses to many data consumers. In addition, data virtualization technology offers a long term benefit by allowing companies to incrementally build up a set of shared data services that can be used (and re-used) in operational and analytical business processes across the enterprise.

Data virtualization can also shorten time to value in data warehousing environments while reducing cost and complexity

Combining data from multiple data sources for access in a self-service BI environment is particularly powerful as it simplifies access to data and therefore allows business users to spend most of there time analysing data rather than trying to integrate it. In addition, we are now at the point time to value is becoming critical in data warehouse environments. Data virtualization offers this capability allowing prototypes to be built quickly with business user involvement from the start even if the data needed is not all in a data warehouse. This is a much more agile approach to data warehouse development. In addition we have also seen that data virtualization simplifies data warehouse architecture by making it possible to create virtual data marts instead of always needing physical data marts. Furthermore, it can support master data management initiatives and support data quality while integrating data on-demand. This kind of flexibility makes data virtualization software a 'must have' technology as part of a data management platform of tools.

SAS has strengthened its position in the market by adding support for data virtualization

With SAS Federation Server as part of the SAS Data Management capabilities, SAS strengthened their position in the marketplace along with their right to be included in any organisations short-list of vendors to supply an end-to-end offering for enterprise data management.

About Intelligent Business Strategies

Intelligent Business Strategies is a research and consulting company whose goal is to help companies understand and exploit new developments in business intelligence, analytical processing and enterprise business integration. Together, these technologies help an organization become an *intelligent business*.

Author

Mike Ferguson is Managing Director of Intelligent Business Strategies Limited. As an analyst and consultant he specialises in business intelligence and enterprise business integration. With over 30 years of IT experience, Mike has consulted for dozens of companies on business intelligence, enterprise architecture, business integration and data management. He has spoken at events all over the world and written numerous articles. Mike is a resident expert on the Business Intelligence Network, providing articles, blogs and his insights on the industry. Formerly he was a principal and co-founder of Codd and Date Europe Limited – the inventors of the Relational Model, a Chief Architect at Teradata on the Teradata DBMS and European Managing Director of Database Associates. He teaches popular master classes in Business Intelligence, Enterprise Data Governance, Master Data Management, and Enterprise Business Integration.

Water Lane, Wilmslow
Cheshire, SK9 5BG
England
Telephone: (+44) 1625 520700
Internet URL: www.intelligentbusiness.biz
E-mail: info@intelligentbusiness.biz