March 2013

The Great Rethink: Building a Highly Responsive and Evolving Enterprise Data Integration Architecture

Claudia Imhoff, Ph.D.
Joe McKendrick

Table of Contents

FIGURES

Introduction

Competing and gaining the edge in today's global 24x7x365 economy requires unfettered, around-the-clock access to the abundance of data flowing through enterprises. The problem is, decision-makers are far from getting this data – the methods their organizations rely on to deliver data are often slow, scattered and fractured.

That's why business decision makers and data professionals need to rethink how they go about transforming data – now emanating from a variety of sources and in a wide array of formats – into information they can use to grow their businesses. To bring data together, organizations have long relied on data integration (DI), which typically has been relegated to a back-office activity of manual coding, patching and connecting.

Now, DI is awakening, and evolving into a multidisciplinary architecture – well beyond its traditional extract, transform and load (ETL) mission, to encompass activities including data quality, data federation, data virtualization, master data management, and data exchange.

"Integration is an architectural pattern," says David Linthicum, noted IT speaker, analyst and author of the seminal work, *Enterprise Application Integration.* "Like any architectural pattern, you can improve and refine integration into something more productive and innovative. Integration is a journey, not a project."[1]

Rather than single feeds tied to specific applications or projects, the always-on enterprise must have a well-designed evolving high-level architecture that continuously provides trusted data originating from a vast and fast-changing range of sources, often with different formats, and within different contexts.

In this thought leadership paper, we will show you how to take the journey to sustainable and repeatable data integration practices by building an enterprise DI architecture.

1 From the foreword to *Lean Integration: An Integration Factory Approach to Business Agility*, John G. Schmidt, David Lyle, Pearson, 2010.

Today, It's JBOD Architecture – Just a Bunch of Data

Many enterprises are not yet data-driven, and are still bound to slow, scattered or fractured data environments. Applications and interfaces have historically been built and operated independently of one another. DI efforts are often treated as one-time, standalone projects that deliver views limited to specific processes, functions or departments, but rarely offer a more expansive view of other parts of the enterprise. "Today, many data integration specialists still build one independent interface at a time—a poor practice that is inherently anti-architectural," says Philip Russom, research director with The Data Warehouse Institute.[2]

This dysfunctional lack of cohesive architecture hinders the productivity and agility of the organization itself. Decision making is hobbled by a JBOD architecture, or "Just a Bunch of Data," in which separate pools of data are stored and managed within multiple systems, and employed in one-off instances by business intelligence or analytics software.

Where DI does occur, it is often too complex, too expensive, too manual, and too slow of a process to bring the right data, at the right time, to today's always-on enterprises. In most organizations, in fact, fewer than one out of 10 decision makers and employees have access to the BI and analytic data they need.[3] Such approaches no longer fit in an era where organizations are turning to big data, real time and cloud solutions for competitive advantage.

The data needed for decision making is drawn from a range of sources, including core systems such as ERP, financial, CRM, supply chain and HR applications, as well as peripheral sources such as productivity applications, social media, and sensors. The traditional means of acquiring key data has been via the extract, transform and load (ETL) processes into target systems such as data warehouses or analytics and reporting platforms. Often, connectors and adapters are added between applications, frequently written by analysts or data professionals on an ad-hoc basis. Because of the time it takes to complete this cycle, decision makers time and again find it faster to extract their own data and load it into spreadsheets.

---

2 Philip Russom, *Next Generation Data Integration*, The Data Warehouse Institute, second quarter 2011.

3 Joe McKendrick, *Opening Up Business Intelligence to the Enterprise,* Unisphere Research, a division of Information Today Inc., October 2012.

ETL and data warehousing have worked well for many years across relational database environments that handled mainly internal transaction data. However, these approaches need to be expanded and enhanced to meet the needs of decision-makers looking for insights from a vastly changed data landscape.

As the more traditional approaches to DI fail to keep pace with expanding data environments and requirements, some vexing challenges are arising, including the following:

High maintenance requirements: To get at datasets needed for analytic applications, data professionals build SQL or XQuery statements that tend to get complex and are specifically customized for the project at hand. As a result, decision makers and their data professional partners need to reinvent the wheel every time they get involved in a new project, hand-coding cleansing rules and transformations. It's often not possible to reuse these integration scripts in follow-up projects.[4]

High costs: Attempting DI against projects relying on extraction methods dependent on one-off queries and ETL is costly. The labor costs associated with manual coding and scripting – still prevalent in most shops – take up a significant amount of time and resources – in fact, the length of time to make changes for new reports or new data can take months. In the meantime, decision makers may need new reports or data in a matter of days or even hours. A study of 122 organizations by David White, senior research analyst with the Aberdeen Group, calculates the average cost of DI and transformation to be $1.1 million a year, including software licenses, service and support, and internal support.[5] On average, enterprises have up to four full-time employees working on DI projects.

In addition, with data stores growing at a clip of up to 30% or more each year for most companies, the solution has been typically to throw more hardware at the problem – companies are upgrading hardware or moving to appliances to scale, which is an expensive approach. In one survey of 611 data managers, two out of three respondents say they normally react to data performance issues by upgrading their server hardware and processors. A majority (53%) upgrade or expand storage systems themselves to counter data growth.[6]

---

4 Ash Parikh, "Going Beyond Looking Under the Hood – Doing Data Virtualization Right," March 2, 2012, http://ashparikh.sys-con.com/node/2185717

5 David White, *Future Integration Needs: Embracing Complex Data*, Aberdeen Group, June 2011.

6 Joe McKendrick, *The Petabyte Challenge: 2011 IOUG Database Growth Survey*, Unisphere Research, a division of Information Today, Inc., and the Independent Oracle Users Group, August 2011.

Latency: It typically takes time to build interfaces to data sources, and the complexity of today's highly varied data environments is only making things worse. In a recent survey of 400 attendees at conferences held by The Data Warehouse Institute, two-thirds of respondents admitted that it took “more than one month” to deliver new data views after a re-organization, and 64 percent said they needed more than a month to integrate a new data source into their data warehouse. These findings suggest organizations are dealing with increasingly complex environments with highly manual processes.[7] Today’s hand-coded, ETL-based systems are not effectively delivering insights to decision makers when and where they are needed. As companies move to using their data warehouse and business intelligence infrastructure for operational decisions, it’s important to have right-time data. However, most data is still more than 24 hours old, which may be too old for operational applications, such as a logistics or customer care systems.

Lack of trust: The inability to provide the right data in a timely manner breeds a lack of confidence in existing data. Decision makers are forced to go elsewhere for the information they need, because current data warehouse environments are often built to serve analytical audiences having unrelated purposes, and don't provide timely information. In addition, decision makers may need data from outside traditional warehouses and databases, such as variably structured data – documents, video, social media interactions. As information moves closer to real-time versus batch mode, the analytics delivered are evolving from being merely “descriptive” to more “prescriptive,” which triggers actions. Trust in data being provided is becoming more critical.

Inability to assimilate new forms of data: Data is becoming varied, beyond the scope of current ETL and data warehouse capabilities. Today's data infrastructures were designed to manage and process structured, relational data. However, today’s business needs require that data that is in variably structured formats – including weblog, social media, document, image, text, and graphical files. Variably structured or non-relational data—including but not limited to video, audio, graphics, web, social media, and office documents—is part of core enterprise information today, in a big way. A majority of 338 data managers participating in one recent survey, 73%, say the volume of variably structured or non-relational data types in their organizations has increased over the past three years.[8] More than a third, 35%,

7 John Evans, “Survey Results on ETL-Based Data Warehouses,” November 1, 2012. http://blog.kalido.com/survey-results-etlbased-data-warehouses/

8 Joe McKendrick, *Moving Data: Charting the Journey from Batch to Blazing*, Unisphere Research, a division of Information Today, Inc., May 2012.

consider this growth to be intense, reporting that the overall volume of this type of data increased by more than 25%.

In addition, at least 25% of the data now being handled through integration efforts is external data. While organizations have some control over internal data, external data is far more difficult to manage, as the sources of this data may be unknown, may not be high quality, and the provider may not be helpful. A recent Aberdeen survey finds only five percent of survey respondents were able to integrate external data without manual intervention.[9]

9 David White, *Future Integration Needs: Embracing Complex Data,* Aberdeen Group, June 2011.

Why A Well-Designed Data Integration Architecture Makes a Difference

It's time to rethink the DI process, and move it toward an agile, responsive enterprise environment. In the information technology world, various forms of architecture – especially enterprise architecture and data architecture – help the business define its requirements and plans. Enterprise architecture focuses on the systems and applications, and data architecture is concerned with databases, data models and semantic models. An enterprise DI architecture builds upon these methodologies, providing a comprehensive framework that supports repeatable DI strategies across all applications and data sources.

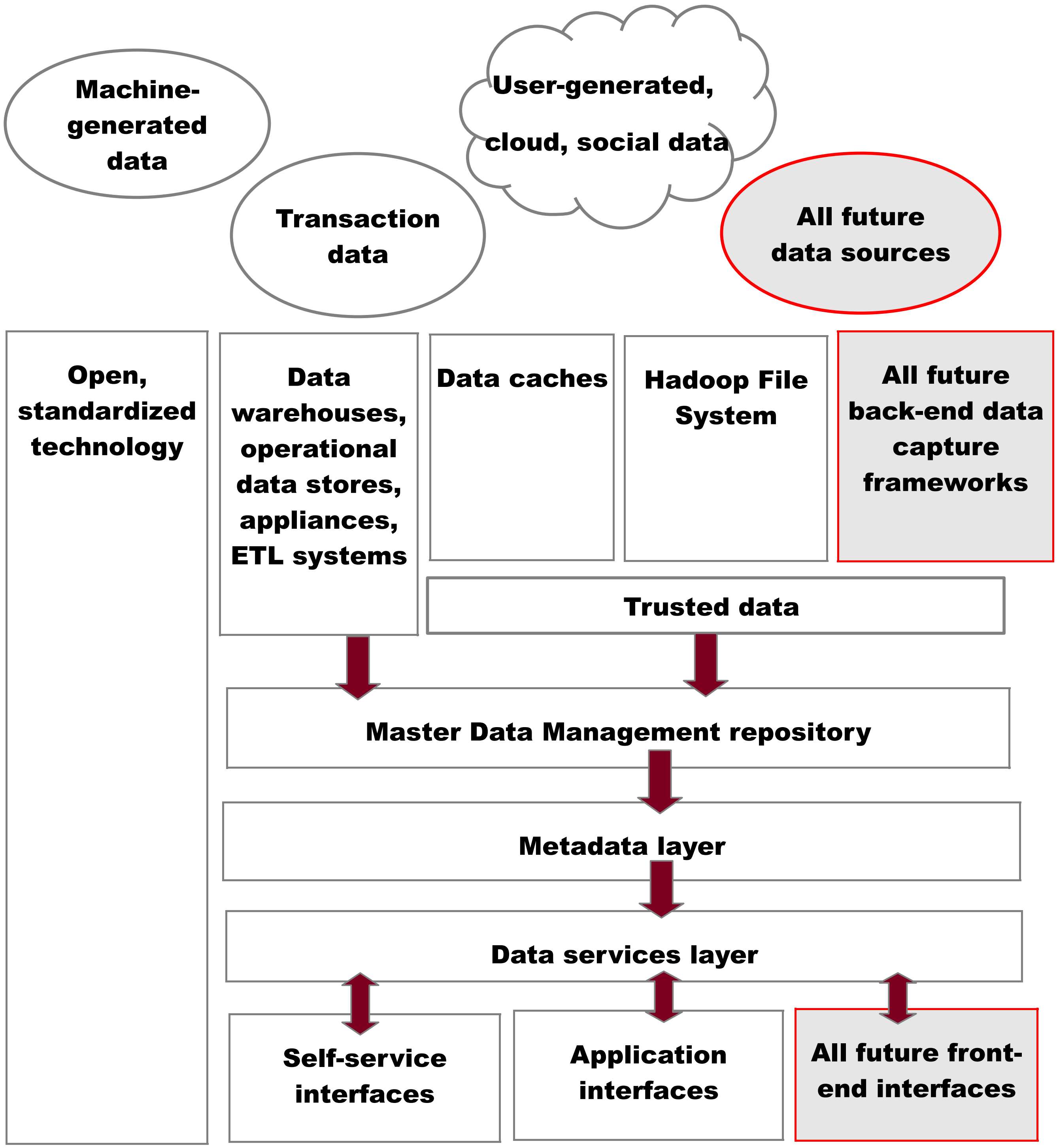
An enterprise DI architecture is the result of an evolution of DI from project-based activities to a strategic initiative. It transforms DI into an ongoing, continual, and repeatable process that delivers value to all parts of the organization. An architectural approach is based on buy-in for DI from all parts of the business, to provide information that can be reused across the enterprise, via automated methodologies.

Enterprise DI architecture is not tied to any one technology or methodology. A well-designed architecture should support existing ETL-based interfaces, a combination of ETL, data federation and virtualization, or some other type of approach. The purpose of a DI architecture is to enable change within the interface and data structures as the business requires, along with the addition or replacement of interfaces and data sources without disrupting the enterprise data fabric.

As shown on the chart on page 9, data on the back end is be captured from a range of sources, from transactional systems, machines and applications, users, cloud or social-media, and any and all other sources that arise in the future. The data may be captured within operational data stores, data warehouses, appliances, or through emerging frameworks such as the open-source Apache Hadoop File System. The data flowing into the enterprise is identified, deduped and cleansed to assure it trusted data that has the confidence of decision makers downstream. Data quality processes are often already in place within existing data warehouse and ETL environments; these practices need to be extended to all other enterprise data inputs as well.

Master data management ensures that there is a single, enterprise version of the data. A metadata layer provides details on all the critical information that is captured and available to decision makers across the organization. Moving toward the front end, a data services layer virtualizes the data that is coming in from various sources – such as production databases, data warehouses, and unstructured data – abstracting it for consumption by any and all clients and device types. Finally a self-service framework on the front end enables the rapid configuration of end-user interfaces – such as dashboards and portals – for insights and consumption. Throughout the process. open and standardized technology enables the development and deployment of a range of solutions.

Enterprise Data Integration Architecture Layers



The Journey to Enterprise Data Integration Architecture

Building and sustaining a well-designed enterprise DI architecture requires a retooling of processes that were developed around fractured, un-integrated and project-based approaches to DI. This isn't going to happen overnight, but will evolve as organizations move from project-based DI to a comprehensive strategy for enabling data access and integration potentially in real time across the enterprise.

This requires a rethinking of integration itself – as "a repeatable ongoing process rather than as a custom one-off project effort," says John Schmidt in *Lean Integration.*[10] "While an integration point between any two systems always looks unique from the perspective of the two systems, if you step back and look at all the information exchanges between systems in an enterprise, what you find is a relatively small number of patterns that are repeated over and over again. The details are different for every instance, but the patterns are not."

To accomplish this, a more holistic approach is needed to the way data is identified, managed and made available to the business. Information needs to be moved out of departmental and functional silos. Business decision-makers – not just data professionals – need to take the lead in identifying what datasets are important to analytical environments. To facilitate this process, data professionals need to embrace repeatable processes through automation.

The chart on the next page shows the continuum of stages an organization goes through as it makes the journey from project-based DI to an enterprise DI architectural approach. As organizations move toward adopting an enterprise DI architecture, the DI process becomes more natural to the organization, a part of "muscle memory" that automatically kicks in as analytical and other data management applications are engaged.

10 Schmidt and Lyle, *Lean Integration*.

Data Integration Continuum	
Project-based	There may be multiple, unconnected DI platforms within the same organization. DI takes place on a project-by-project basis, most often through hand-coded scripts, or connectors from individual applications. Knowledge about projects is retained in individuals' heads, versus being documented or systematized as a process. All or most projects take place within departments or business units, and data is not shared across the enterprise. The quality of data delivered is questionable.
Application-centric	Hand-coded scripts are still prevalent, but some DI is facilitated through connectors and interfaces to enterprise applications. Most integration scripting is documented. Data warehouses are employed to capture selected datasets for analysis by selected parts of the business. Most data is still in silos, attached with specific applications.
Process-centric	Some automation of DI processes is introduced. DI expands to cover multiple disciplines, including data quality and data exchange. Some reuse of DI scripts takes place. Metadata and master data management is employed against most core systems to provide a single version of the data. Much historical data is maintained and accessed through data warehouses, and some operational data as well. Some data professionals may also be working with frameworks such as Hadoop.
Architecture-based	DI is continuous and available across the enterprise, via standardized and reusable data services and metadata. Most, if not all, interactions are automated. Data sources are added to the flow as quickly as they are needed by the business. Data is trustworthy, deduped and verified from original sources. Business users can design their own queries and interfaces and access enterprise data with minimal intervention from the IT or data management department. Data is viewed and analyzed on a real-time or right-time basis where required. Lean integration principles – emphasizing repeatability, continuous improvement and quality – are practiced.

12 Building Blocks of an Enterprise Data Integration Architecture

Making the journey from scattered, project-based DI approaches to a holistic enterprise DI architecture requires adopting the right tools, practices and methodologies along the way. The following are 12 key building blocks that will help you make this journey:

1. Put the Business in Charge of Data

The object of any architectural endeavor is to build a bridge between the business and the technology side of the organization – to encourage and establish greater collaboration between business decision makers and data professionals to achieve DI goals. To be successful, an enterprise DI architecture needs to be driven by the business, for the business.

A good place to start is to identify the business executive sponsor for enterprise DI efforts. This person has the ear of C-level executives, and can help line up support across the enterprise. An enterprise DI architecture can be advanced by committees, centers of excellence, or competency centers that can elevate the discussion and planning above and beyond any organizational political issues or turf battles that may hamper the effort. It's also essential to build in and monitor key performance indicators for critical business requirements not being met by the current infrastructure. In addition, organizations need a way to determine what data must be governed and what data may remain outside of governance.

2. Assure the Trustworthiness of Data

A key objective of the data governance effort is to assure the trustworthiness of data. The foundation of data integration architecture is built upon trust – business decision makers need to be confident that the information they are receiving is timely, accurate and represents a single version of the data. One survey of 193 companies, conducted a few years back, found only four percent considered the quality to be “excellent.” At the time, only 37% had a data quality initiative in place.[11] A data quality initiative is based on an understanding of data sources and databases from which data originates, cleaning or “scrubbing” current data, identifying and removing duplicate records, rationalizing data to consolidate records, and securing the data.

---

11 *The State of Data Quality Today: An Information Difference Research Study*, July 2009. http://www.pbsoftware.eu/uk/files/download/analyst-reports/TheStateofDataQualityTodaybyInformationDifference.pdf

3. Get Everyone on the Same Page with Standardization

To get everyone on the same page, enterprise DI architecture proponents need to establish a common set of data and technology standards for use across the enterprise. Using web services-compliant interfaces and data exchanges enable access on both sides of the firewall. Best practices reduce reliance on documents and spreadsheets and increase the use of enterprise data sources. To standardize DI across the enterprise, it's important to establish common definitions and protocols.

It's also important to make the reuse of data services and artifacts a goal. This will simplify multiple DI efforts, and help them more closely adhere to standards and patterns, as well as enable consistency. “The recommended way to promote standards is to make them easy to find, understand, and use,” say Schmidt and Lyle. “To some degree, it is possible to force people to use standards, but a more effective and sustainable approach is to pull everyone to the standards by making them easy to follow.” Standards need to apply to areas as varied as integration tool selection, development lifecycle processes, naming conventions, canonical data models, integration patterns, and security standards.[12]

4. Position Data as a Service

To make data widely available across the enterprise, one of the most promising approaches is enabling data and data management functions to be abstracted or virtualized from original sources via a logical – or virtualized – data services layer, so it is accessible from any type of system, application or device. Data virtualization is foundational to DI, since it enables fast and direct access to critical data and reports.[13] It is a form of service-oriented architecture or internal cloud, meaning important data management tools, software and datasets are exposed as standardized, reusable services.

Information coming from enterprise data sources must be made available to all clients regardless of device or location – whether it's a PC, tablet, smartphone or other device. This reduces the number of interfaces, provides a single, consistent architecture more easily understood by the business, and enables greater productivity, standardization and reuse.

5. Adopt Lean Integration Principles

---

12 Schmidt and Lyle, *Lean Integration*.

13 Ash Parikh, “What it Takes to Be a Leader in Data Virtualization,” January 23, 2012. http://blogs.informatica.com/perspectives/2012/01/23/what-it-takes-to-be-a-leader-in-data-virtualization/

Lean integration – based on the same principles honed over the years in lean manufacturing – emphasizes continuous improvement and the elimination of waste in end-to-end DI and process integration activities.[14] By embracing lean integration, organizations recognize that DI is an evolutionary process that needs to be incorporated into all information technology activities. The seven principles of lean integration include 1) focus on the customer and elimination of waste; 2) automate processes; 3) continuously improve; 4) empower the team; 5) build quality in; 6) plan for change; 7) optimize the whole.

6. Recognize That DI Has a Lifecycle

An enterprise DI architecture is not launched in one “big bang,” but rather, is an evolving process. Every aspect of data integration work moves through a lifecycle, moving through the three stages of initiate, invest and evolve. Elements of DI – such as ETL, Hadoop, self-service or the addition of new solutions – will be at different stages of their lifecycles, maturing from original initiation to continuing investment in the process. Finally, the process or practice evolves to the point where it either needs to be retired or transformed, and this calls for a return to the initiate stage.

7. Automate DI

There are two levels of productivity that are reshaped as part of the move to an enterprise DI architecture. First are the productivity requirements of the business end-users. Often, these users need information on a moment's notice, as they respond to changing situations like new market demands or customer requirements.

Second, there is the productivity of IT and data management departments that need to turn around information and reports as fast as possible, while still maintaining operations. The most important consideration in effective DI is to reduce reliance on manual integration methods, such as hand coding or scripting. It's also important to adopt new approaches such as data virtualization and data propagation (e.g., message brokering). In the Aberdeen Group study of 122 companies, 83% of those employing automated DI tools were able to deliver data on a timely basis, versus 33% using manual methods.15[15] “Organizations that are able to automate the process of updating data are far more successful at getting information into the hands of managers in the timeframe required,” says Aberdeen. Other techniques include moving to graphical or visual DI tools with pre-packaged functionality, and adopting data quality tools including discovery, profiling and cleansing solutions. Repetitive DI tasks – such

14 Schmidt and Lyle, *Lean Integration*.

15 White, *Future Integration Needs*.

as testing and monitoring – can be automated, thereby saving countless hours of manual reviews.

8. Incorporate Metadata and Master Data Management into the Enterprise

Through metadata and master data management, the enterprise has a core set of descriptive data and reference data. These two distinct sets of data have typically been buried within IT and data management departments, but will be a key part of enterprise DI architectural approaches as they emerge and are brought closer to the business.

Metadata – a centralized repository of data about data – should capture and represent the most critical data from across the enterprise. Metadata describes both operational and decision support data, and consists of information about systems, processes, and information and how they work together. Metadata tracks the lineage from start to finish of data and assigns value to information maintained within enterprises.

Master data management allows for a single "gold copy" of reference data pertaining to core business entities (people, products, places). In the enterprise there are several systems managing the same data; company acquisitions or mergers are another example when an enterprise gets several parallel systems managing similar and sometimes overlapping data. The role of MDM is to centralize the data management to only one master copy of the data item that is then synchronized to all applications using that data. Using this approach means that all systems refer to the same customer, not fractured subsets of the customer information.

9. Move to Self-Service

Self-service interfaces will enable decision makers to design their own queries and quickly access information and answer queries without the need to go through their IT or data management departments. The key is to develop and provide interfaces that are highly intuitive and simple to use for business decision makers. The interface shouldn't resemble the instrument panel of a 747; it should have the look and feel of the dashboard of a car.

A good place to start is to identify areas of the business that require rapid access to key metrics and analytical results. It is possible to replace or shorten the requirements gathering process with a self-service one that allows rapid prototyping. The logical layer can be built up project-by-project, which requires planning ahead so that future projects can reuse the work in current projects. It's important to note that IT will still maintain a role in self-service, managing processes, reusing prototypes and security, monitoring overall activities. Self-service BI does not mean self-sufficiency or BI chaos!

The challenge is enabling self-service systems to reach a wider audience of users than previous, and more constrained, BI systems have. Self-service BI can provide new insights in areas such as financial analysis, forecasting, and business activity monitoring. A self-service data and analytics platform enables the decision-making process with a simple user interface, and exposes data by making it available to people who need it. Decision-makers can configure their own reports, develop their own analytics, and share their results with others – all without significant IT involvement – thus enabling faster and more-timely decisions.

10. Open Up Platforms and Technologies

Enterprise DI architecture is not wed to a single type of technology solution, platform, or even standard. Rather, it is inclusive of existing solutions, while being open for any new technologies or solutions that are incorporated later. An enterprise DI architecture accommodates both open source and commercial solutions that allow for unlimited data growth. The solutions also handle all data types – traditionally structured and new, variably structured sources.

There isn't one right answer for the types of data structures needed to support an enterprise DI architecture, and technology keeps changing all the time. An architecture supports the ability to develop and nurture a "co-existence" strategy that brings together multiple and diverse data environments. Newer approaches or frameworks such as Hadoop can be blended into existing technologies such as relational databases, data warehouses, data marts, or operational systems—and any and all other platforms that are best-suited to deliver the analytics capabilities end users require. A survey of 298 data managers conducted among members of the Independent Oracle Users Group finds that one-third have developed ways to pre-process Big Data, then load it into their data warehouse for integrated analysis. This suggests that the majority of respondents who are using Big Data find the greatest value in integrating the emerging variably structured data world with existing relational data environments.[16]

11. Strive for Scalability

An enterprise DI architecture needs to grow with the organization as it generates and consumes greater volumes and varieties of data. The availability of open-source frameworks such as Hadoop enables processing and analytics across variably structured data files deemed too costly or cumbersome to manage within relational or data warehouse environments. Hadoop can be employed to augment

16 Joe McKendrick, *Big Data, Big Challenges, Big Opportunities,* Unisphere Research, a division of Information Today Inc., in partnership with the Independent Oracle Users Group, September 2012.

existing BI systems in an extended environment. This will enable additional processing and analysis of large volumes of this new form of data. Hadoop can be integrated into relational/data warehouse capabilities into a single view across diverse data sources. Workloads can be offloaded from more expensive data warehouses into experimental marts or “sandboxes” for separate analysis.

In addition, a successful DI architecture enables a range of strategies that allow for growth in the data environment – such as grid computing, in which workloads are distributed across different computing nodes across a network – will help better manage peaks in workloads. Additional scalability strategies that can be addressed within a DI architectural framework include the deployment of standardized hardware, such as commodity server boxes or blade servers. It's critical that database and systems administrators be able to quickly analyze data usage to determine peak workloads, determine where bottlenecks may be occurring, and resolve those bottlenecks. In addition, part of this scalability will be getting ready for Hadoop by making sure the existing infrastructure will support it.

12. Move to Real-Time and Right Time

The ability to move data rapidly from sources to decision makers or applications is another hallmark of a successful enterprise DI architecture. To enable real-time or right-time delivery of data, a tiered cache approach should be employed. In addition, it may not be possible to put everything in the data warehouse. The data warehouse by its nature must have some level of data latency in the integration of its stored data. To increase real-time visibility and analytics, you may need to combine real-time DI with batch DI processes.

Relational databases and data warehouses are still the primary platforms by which most organizations deliver historical analytical data. In these environments, operational data stores can be deployed as caches for more current operational decision-making. In addition, high-volume sites are shifting to newer types of solutions, such as appliances and data virtualization to move data faster, a survey of 338 data managers shows.[17] Respondents recognize the advantages cloud computing can bring to real-time DI, and many efforts are now underway, mainly with private or hybrid cloud-based services. Most already have some workloads and data in the cloud, and expect this to keep growing.

Of course, if real-time analytics are required on real-time data, our traditional BI solutions must be extended to embrace stream analytics – that is, analysis of the data as it is streaming into the enterprise.

17 McKendrick, *Moving Data: Charting the Journey from Batch to Blazing*.

Finally it is important to understand the true requirements for data latency. The business may be confusing real time access with real time data. IT implementers and business end-users must determine what data must be real-time versus low latency or historical data. Then it becomes possible to create combined views from these different latencies, via real-time caches or data virtualization.

This new architecture, embracing traditional data warehouses and data marts, experimental or sandbox technologies, and finally streaming analytics on real-time data is what we refer to as the extended data warehouse environment.

Summary

In many organizations today, productivity and agility are hindered by a dysfunctional lack of cohesive architecture – a JBOD architecture, or "Just a Bunch of Data," in which DI is only employed on a project-by-project basis. This results in high maintenance requirements, high costs, high latency, lack of trust in data, and an inability to assimilate new data sources and formats.

The goal of an enterprise DI architecture is to deliver a data environment uninhibited by these functional and departmental silos in order to serve the enterprise and overcome the previous dysfunctions. Building and sustaining a well-designed enterprise DI architecture requires a retooling of processes that were developed around fractured and project-based approaches to DI. This isn't going to happen overnight, but will evolve as organizations move from project-based DI to a comprehensive strategy for enabling data access quickly and in real time across the enterprise.

This architectural approach to DI improves the capacity, productivity, timeliness and value of information and incorporates existing tools and platforms as well as next-generation frameworks.