IBM

**The dynamic warehousing
infrastructure:
Establishing a foundation to meet
new information requirements**

Introduction

Traditional data warehouses are increasingly being challenged by demands for real-time data access, analysis of structured and unstructured data, and the need to synchronize core customer and product information across operational systems to create a single view of the enterprise. These changes are the result of new business requirements to leverage enterprise information more effectively in order to:

- *Identify new opportunities and deliver new products to market faster.*
- *Optimize business processes through real-time information and analytics.*
- *Provide increased visibility to business performance.*
- *Meet industry compliance standards for reporting.*

Yet most long-standing data warehouses are designed to support a relatively small number of users who access information to support strategic decisions, financial planning and the production of standard reports that track performance. Today, many more users need to access information in context and in-line so that critical functions are optimized to run efficiently. Information about customers—both structured and unstructured—must be analyzed and delivered wherever it is needed. Key performance indicators (KPIs) should also be available at all times to monitor performance. In short, business intelligence is becoming embedded in key business processes.

To create a true enterprise view of information that supports strategic and operational functions, enterprise data warehouses must be reinvented as a dynamic source of current and historical information. Dynamic warehousing is an approach that enables organizations to deliver more dynamic business insights by integrating, transforming, harvesting and analyzing insights from structured and unstructured information. Capable of processing large amounts of information, a dynamic warehousing infrastructure can enable organizations to respond on demand to unscheduled analysis requests, and as events trigger the need for information throughout the day.

A key component of a dynamic warehousing environment is the data warehouse platform. To implement a dynamic warehouse, the platform should be able to:

- *Process transactions and analytical requests.*
- *Handle varying service level agreements (SLA).*
- *Scale easily as the number of applications grows.*
- *Analyze structured as well as unstructured data.*
- *Provide real-time analytics that can be embedded in business processes.*
- *Support advanced analytics such as data mining within the data warehouse.*

But the requirements for dynamic warehousing go well beyond having the right data warehouse platform. Dynamic warehousing requires an extended infrastructure (Figure 1) to:

- *Implement changes to the business model without impacting usage.*
- *Monitor and analyze data sources for structure and content to ensure the best data is being accessed for each application.*
- *Provide tools that enable business users and IT staff to collaborate on data requirements and definitions.*
- *Deliver impact analysis and data lineage reports to coordinate changes and provide visibility to critical data flows.*
- *Deliver data that has been cleansed and harmonized to the warehouse regardless of volume or latency requirements.*
- *Synchronize master data for key business entities across operational systems.*
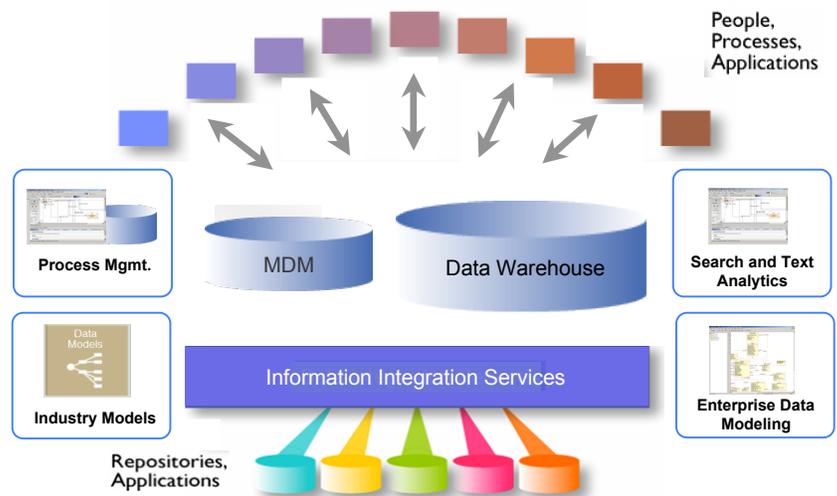

Figure 1: Dynamic warehousing has an extended infrastructure that leverages metadata to extend the data warehouse to operational systems as well as traditional business intelligence applications.

This white paper outlines the options and decision criteria for selecting the right components that define the extended infrastructure for dynamic warehousing, including how to:

- *Access and integrate source system data.*
- *Gather requirements and implement changes to the data warehouse model.*
- *Synchronize master data for key business entities across operational systems.*

By creating a roadmap for a truly dynamic warehousing infrastructure, organizations can meet their most pressing needs for business intelligence today, while ensuring a data warehouse environment that can help support rapid growth, significant change and increasing demand for real-time information.

The industry data model: Foundation for dynamic warehousing
A critical component of any data warehouse infrastructure is the data model that specifies how information is structured and how it is accessed for analysis and reporting. Traditional data warehouses are often built upon hand-coded or application-specific data models that fall short of requirements for dynamic warehousing. For example, these data models do not usually incorporate industry best practices, compliance reporting or enterprise-wide reporting perspective. Over time, the models often become fragmented, costly and increasingly difficult to use, as well as slow to respond to changes in business requirements. This makes comparability across reporting solutions inconsistent and therefore useless for effective decision making.

A better approach is to use an industry-specific data model that incorporates best practices and includes structures that are common and well understood by your organization. A data warehouse built using an industry data model encapsulates extensive best practices in delivering effective data warehouse solutions. Tailored to address the specific needs of an industry, the integrated, interlinked and customizable industry data models contain thousands of hours of development efforts and expertise to help business users and IT implement business-ready analysis templates (business solution templates) and an enterprise data warehouse on time and on budget.

In addition, an effective industry data model can give organizations the flexibility and scalability to start with a small implementation and easily add more business solutions as they are needed, with minimal disruption to ongoing usage.

Overall, an industry data model should incorporate three critical functions:

Identify and define requirements through collaboration

An effective data model enables business users and analysts to easily scope reporting and analysis requirements, including identifying the KPIs and the required data inventory. To help ensure that all interested business and IT parties are working from the same information, an industry-specific glossary of terms is needed to aid in the recognition and understanding of requirements and downstream reporting structures.

Called an *industry business terminology data model*, this glossary functions as an enterprise-wide vocabulary of business concepts that can be used to gain agreement across all aspects of business and IT on how a company defines itself and its measurements for success, such as profitability. This information is organized into a customizable nine element hierarchy of business terms and definitions that are directly linked to other areas of the model, including the KPIs, and the underlying data definitions found in the enterprise data warehouse. Designed to accelerate the model development process, the industry business terminology data model can help maximize the value of existing information by enabling business and IT to speak the same language during requirements definitions and implementation phases of a project.

Establish a single, unified view

During the data integration process, the industry data model becomes the consolidation focal point. It not only supports the harmonization of data from disparate information silos, but also becomes the data source for downstream reporting data marts that are accessed by business intelligence tools. By bringing data together from across various disciplines, the enterprise data warehouse model (design model) defines a single analytical view, or single data repository, that is used to ensure a broad range of applications and analysis can be supported and new functions added without having to completely rebuild the data model.

Provide access to pretested business solution templates

Industry business solution templates (requirements model) are a collection of business-centric KPIs that make up the enterprise's consolidated business reporting metrics and represent industry, financial, risk and compliance reporting best practices. For example, a banking data model might include multiple, pre-built analysis that focuses on areas like relationship marketing or lead analysis, which is made up of measures such as customer lifetime value and dimensions such as campaign type and competitive win status (see Figure 2).

Industry Data Models – Business Solution Templates

Banking (Banking Data Warehouse)	Financial Markets (Financial Markets Data Warehouse)	Health Plan (Health Plan Data Warehouse)
▪ Profitability	▪ Risk Management	▪ Claims
▪ Relationship Marketing	▪ Asset and Liability Mgmt	▪ Medical Management
▪ Risk Management	▪ Compliance	▪ Provider and Network
▪ Asset and Liability Mgmt		▪ Sales, Marketing and Membership
▪ Compliance		▪ Financials
Insurance (Insurance Information Warehouse)	**Retail** (Retail Data Warehouse)	**Telco** (Telecommunications Data Warehouse)
▪ Customer centricity	▪ Customer centricity	▪ Churn Management
▪ Claims	▪ Merchandising Management	▪ Relationship Mgmt & Segmentation
▪ Intermediary Performance	▪ Store Operations & Product Mgmt	▪ Sales and Marketing
▪ Compliance	▪ Supply Chain Management	▪ Service Quality & Product Lifecycle
▪ Risk Management	▪ Compliance	▪ Usage Profile

Figure 2: An effective industry data model includes business solution templates that can help generate immediate business value.

These templates are not only used to help business managers quickly identify the scope and customize their analytical reporting requirements, but they can then be used by IT to implement the data marts and OLAP cubes used by dashboard, scorecard or ad hoc reporting tools to gain business insight.

Ensuring trusted information: Enterprise data integration
Many organizations have begun to examine ways to integrate enterprise data to help derive more value from the complex, heterogeneous information spread across their systems.

A true enterprise platform for data integration should deliver all the functions required to integrate, enrich and deliver information that you can trust for your key business initiatives. To provide a strong foundation for enterprise information architecture, the platform should include rich functionality, broad connectivity to heterogeneous sources and a unified, metadata-driven approach.

Typically, however, organizations design and deploy data integration solutions for a single business application at a time. With each new system or project, the data integration efforts begin anew. In addition, there is also often a lack of standards or best practices in place to ensure that the information resulting from the data integration effort meets end-user expectations — and no viable processes to track the quality of the information from source systems over time.

Over time, the complexity can deepen as the amount of data collected and processed continues to grow dramatically. Completely customized integration of all the data for each of an organization's myriad systems and silos requires new tools for each integration project and carries a potentially crippling productivity penalty. What's more, completely customized integrations rarely provide the positive bottom-line contribution needed to allow the organization to remain agile and competitive.

In order to create a dynamic warehouse environment that delivers trusted information consistently and reliably, organizations must adopt a unified platform for data integration that is capable of supporting every step in a data integration life cycle, including:

Data profiling

A unified platform for data integration can help you understand and analyze the meaning, relationships and lineage of information. By automating data profiling and data quality auditing within systems, you can help:

- *Create an understanding of data sources and relationships.*
- *Eliminate the risk of using or proliferating inaccurate or inconsistent data.*
- *Improve productivity.*
- *Leverage existing IT investments while increasing capabilities.*

Data cleansing

Data cleansing capabilities help ensure auditable data quality and consistency by standardizing, validating, matching and merging information to create comprehensive and authoritative information for multiple uses. The data quality logic should be deployed universally, ensuring data consistency and accuracy across the enterprise. By improving the quality of information, your organization can help:

- *Make more efficient and effective business decisions with trustworthy data.*
- *Reduce IT and supply chain costs with a single, accurate view of customers, products and suppliers.*
- *Improve customer service and revenue-generating opportunities.*
- *Assemble the auditable, trusted information needed to comply with regulations such as the Sarbanes-Oxley Act and Basel II.*

Data transformation

An enterprise data integration platform must also transform and enrich information for new uses in new contexts. Hundreds of pre-built, metadata-driven transformation functions combine restructure and aggregate information from its current application-centric form into entirely new contexts, allowing information to be used in new ways to suit new business needs. Using transformation, you can help:

- *Remove the complexity of integrating data from heterogeneous data sources.*
- *Derive important and relevant information out of complex, heterogeneous data.*
- *Ensure information is in a form appropriate for its intended use.*
- *Provide an enterprise-wide view of the business at any time, to anyone.*

Data delivery

With the ability to virtualize, synchronize or move information to deliver it in-line to the people, processes or applications that need it, a unified platform for integration provides information when you need it and how you need it. Whether the information is delivered on demand, through federation or on a timed or event-driven basis, it can be moved in bulk from location to location or accessed in place reusing the same core logic. Information delivery helps enable you to:

- *Ensure information is always available, when and where it's needed.*
- *Improve data accessibility and consistency for self-service operations.*
- *Reduce latency for real-time visibility into operational information.*

Metadata is the cornerstone

Another critical component to a unified data integration platform is a common metadata repository that links with the data modeling tools and industry data models (see Figure 3).

A unified, metadata-driven infrastructure facilitates a shared understanding across business and technical domains while helping you reduce the time between specification and build. Furthermore, it can provide a transparent and persistent record of understanding that can drastically reduce future project delivery times and improve overall insight and confidence in information.

An effective, metadata-driven enterprise data integration solution uses the three primary types of metadata typically generated by organizations:

- *Operational metadata provides authoring information, creation date, physical location and other operations-based information.*
- *Business metadata identifies the business processes and analytical applications to which raw data should be mapped. Business metadata is critical for end users – or consumers – of information so they can be confident that the data they rely on for making business decisions is exactly what they expected.*
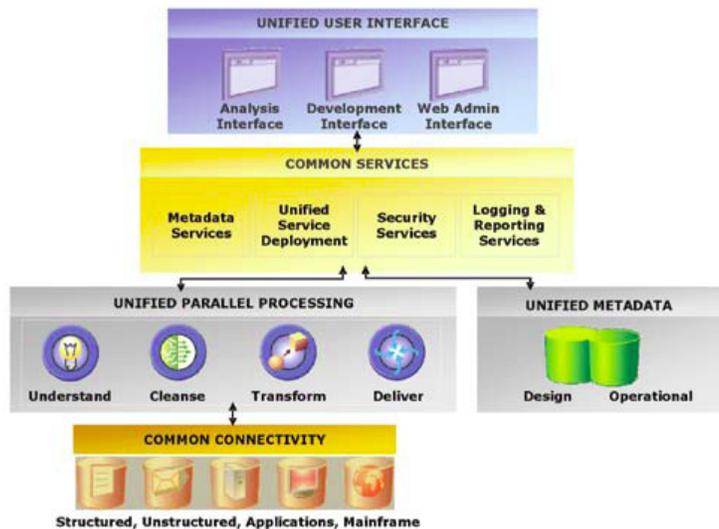

Figure 3: A unified platform for enterprise data integration.

- *Technical metadata provides information on the applications and systems that are being used to store, verify, aggregate and cleanse raw data. It helps companies understand what information they have today and the reliability of that information, and streamlines development efforts by providing technical users with information about the data elements and how they are implemented currently across various systems. Additionally, this metadata can help establish a metadata map of source systems that can be leveraged to build federation queries, improve audit capability and provide visibility through impact analysis and direct lineage reporting.*

Enable information integration services

A Service Oriented Architecture (SOA) takes all of the metadata, methods and processes inherent in a good enterprise data integration platform and makes them work at their most effective and efficient level. A SOA is capable of working with custom, third-party and legacy systems to extract and deliver data through the ability to turn the integration processes, methods and tools into "services" that can be used and reused for virtually every system, every application, and every type of data a business creates. In short, SOA is the enabler that can transform yesterday's incremental gains into tomorrow's data integration breakthrough.

Extend the value of dynamic warehousing with master data

Dynamic warehousing can also play a critical role in developing and using master data for the enterprise. A master data management (MDM) solution gathers and manages the key or common information for core business entities like customer, product and location. It enables the owners of record for different types of data — product managers, for instance, or finance managers — to manage the state, context, and definitions of their master data such as clients, products, or locations. At the same time, operational systems and applications leverage the MDM system as a single operational view of the master data in order to verify the accuracy and consistency of master data used in routine transactions such account opening. Finally, users of the dynamic warehousing can benefit from the reference source of master data fed into the data warehouse.

A dynamic warehouse provides the key metrics and analytics required by business processes and will often feed some of this information to the Master Data Management solution. To accelerate implementation, the MDM solutions should share common data definitions with the industry data models for key party elements such as customer, thereby simplifying the integration and exchange of data.

Accelerate the value of information to Information On Demand

Having the right data warehouse platform can help ensure that a warehouse is dynamic and can thus handle the new requirements for real-time analytics and analysis of unstructured data. **IBM DB2 Warehouse** is a complete enterprise data warehouse offering that provides a best of both worlds architecture, which has been optimized for both transactional and analytical queries required for dynamic optimization of business processes. It also provides embedded analytic capabilities required for next-generation business intelligence, including in-line analytics, real-time data mining, unstructured data analysis and OLAP-based cubing services. DB2 Warehouse is fully integrated with the IBM Industry Models and IBM Information Server, and can leverage the master information managed by IBM Master Data Management Server.

The **IBM Information Server** is a unified platform for enterprise data integration that serves as a buffer between operational systems and the dynamic warehousing environment, helping to create a constant and reliable flow of cleansed information to the warehouse while ensuring that current source systems can be monitored overtime and new sources can be incorporated easily. Built on a unified metadata infrastructure and working in concert with revolutionary data integration methodology, Information Server helps organizations derive more value from complex, heterogeneous information spread across their systems. With breakthrough productivity and performance for cleansing, transforming and moving this information consistently and securely throughout your enterprise, Information Server creates a unified platform that enables you to access and use information in new ways.

In addition, using an industry data model can speed requirements gathering and provide a rich analytical environment for the data warehouse and ensure that it can be implemented in phases as required and take advantage of industry best practices. **IBM Industry Data Models** encapsulate extensive experience in delivering effective data warehouse solutions to provide a blueprint to help bring data from various disciplines together for a single, unified view.

IBM Master Data Management Server (MDM) software decouples master information from disparate applications and unifies it, helping to ensure consistent, up-to-date information across business processes and transactional and analytical systems, delivering a single view of the truth. IBM Master Data Management Server solutions can also work in tandem with the data models by sharing the same definitions to help simplify the integration and exchange of data.

What to look for:
Functionality requirements

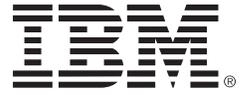| Information integration platform requirements |
| --- |
| ✔ Implement changes to the business model without impacting usage. |
| ✔ Monitor and analyze data sources for structure and content to ensure the best data is being accessed for each application. |
| ✔ Provide tools that enable business users and IT staff to collaborate on data requirements and definitions. |
| ✔ Deliver impact analysis and data lineage reports to coordinate changes and provide visibility to critical data flows. |
| ✔ Deliver data that has been cleansed and harmonized to the warehouse regardless of volume or latency requirements. |
| ✔ Synchronize master data for key business entities across operational systems. |

Data warehouse platform requirements
✔ Process transactions and analytical requests.
✔ Handle varying service level agreements (SLA).
✔ Scale easily as the number of applications grows.
✔ Analyze structured as well as unstructured data.
✔ Provide real-time analytics that can be embedded in business processes.
✔ Support advanced analytics such as data mining within the data warehouse.

Summary

Combining the right warehouse platform and a comprehensive industry data model with a unified platform for data integration that all share business concepts, transformation rules and metadata can enable the deep collaboration between business analysts and IT that is required to deliver information on demand. Add to this a data warehouse platform that is capable of handling the mixed workloads, real-time analysis and advanced analytics, and organizations can deliver a truly dynamic warehousing infrastructure that delivers the right information, whenever it is required, wherever it is needed.

For more information

To learn more about implementing an extended infrastructure for dynamic warehousing, contact your IBM sales representative, or visit: **ibm.com**/software/data/ips/solutions/ddw.html or contact your IBM representative.

IBM.®

IBM.®