EAI and Beyond: A Multilevel Flow Model

By Nigel Stokes, President and CEO,
DataMirror Corporation

DataMirror®

Integrate your data. Anywhere. Anytime.

EAI and Beyond: A Multilevel Flow Model

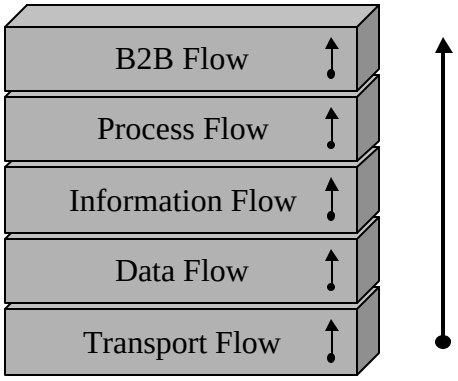
By Nigel Stokes, President and CEO, DataMirror Corporation

Executive Summary

In today's fiercely competitive economy, companies are realizing that new initiatives such as e-Business, customer relationship management and business intelligence go hand-in-hand with a proven organization-wide enterprise application integration (EAI) strategy. The goal of EAI is to integrate and streamline business processes across different applications and business units while allowing employees, decision makers and business partners to readily access corporate and customer data no matter where it resides. More and more, EAI involves integrating information and processes not only across the enterprise but also beyond organizational walls to encompass business-to-business integration.

Companies soon come to realize that an effective EAI strategy involves solving integration problems at several different flow levels: transport flow, data flow, information flow, process flow and business-to-business flow. By flow we mean the delivery of meaningful information to subscribed systems and users in a smooth, continuous stream of bits of information in real-time or near real-time as opposed to the bulk loading of large batches of data at selected intervals.

Multilevel Flow Model - EAI



Caption: Enterprise application integration involves integration at several different levels: transport flow, data flow, information flow, process flow and business-to-business flow.

In their search for an EAI solution, companies are confronted with a vast number of vendors and tools—each claiming in their marketing materials to offer a total EAI solution set. A true multilevel EAI strategy

will incorporate a number of these complementary vendor tools and solutions including software, hardware and consulting to solve integration issues at all of the flow levels.

Resiliency is another EAI imperative. Because the EAI infrastructure touches all aspects of a company's day-to-day operations, it is mission-critical. Companies must have a disaster recovery solution in place to keep all enterprise data and applications highly available and switchable in the event of either planned or unplanned outages. Ideally, the entire EAI infrastructure is mirrored to a recovery server with a mechanism for high speed operational switchover to ensure uninterrupted business operations.

This article will briefly consider the need for enterprise application integration, the various flow levels involved in EAI, specific integration challenges of each level, and technology options available for solving these data integration and resiliency challenges.

The Need For EAI

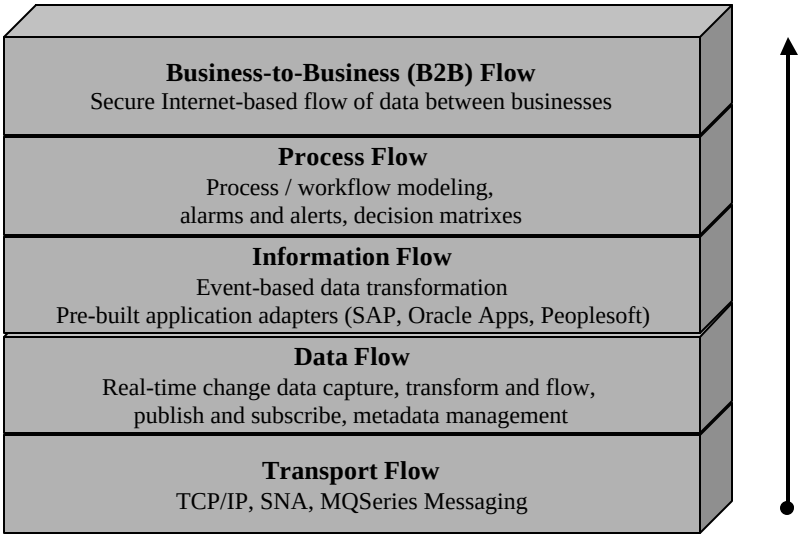
In order to achieve 21st century competitiveness, an organization's core business applications such as enterprise resource planning (ERP), supply chain management (SCM) and legacy systems must be seamlessly integrated with customer relationship management (CRM) and web-based portal applications that provide outward-facing connectivity to suppliers, partners and customers.

In the case of most large organizations, the predominant problem is that systems have been developed or acquired separately at different times from different vendors to support different applications. These systems might have been introduced as a result of mergers or acquisitions or bolted on in haste to handle e-Business or Internet-based trading. Overworked IS departments are frequently being asked to make old systems do something different, and are perpetually finding new and often manual ways to make different systems talk to each other.

To further complicate matters, the systems that drive the business often operate on different hardware platforms and/or database technologies that are incompatible with one another. Adding new applications such as e-Business and CRM to the overall solution set can require considerable effort to integrate with the existing data infrastructure. Application interfaces and collection routines can sometimes grow more complex than the applications themselves.

The overall effect of this complexity is application chaos. Business processes are hindered with resource-intensive manual intervention often needed to move information between key systems. This may involve manually creating extract files or tapes. At worst, it may entail printing information from one system and then re-keying it into another. The challenge of enterprise application integration is to ensure the automatic flow of information between disparate systems across the organization without manual intervention. The EAI methods employed should function transparently without affecting the operation of the participating systems or the application software which runs them. Moreover, the total EAI solution must simultaneously address integration issues at several hierarchical flow levels including: transport flow, data flow, information flow, process flow and business-to-business flow.

Multilevel Flow Model - Components



Caption: EAI Multilevel Flow Model.
Hierarchical flow levels and key components of each level.

Multilevel EAI – Flow Layers

Transport Flow

This foundation layer involves the communications protocols and standards-based technologies employed to inter-network heterogeneous hardware platforms, operating systems and applications. It includes TCP/IP (Transmission Control Protocol/Internet Protocol) which has become the global standard for communications and the routable protocol upon which the Internet is based. Companies seeking to integrate data residing on mainframe systems must also ensure transport flow support via IBM's mainframe network standard SNA (Systems Network Architecture).

The Transport Flow layer also includes messaging middleware that provides an interface between applications, allowing them to send data back and forth to each other asynchronously. Data sent by one program can be stored in a message queue and then forwarded to the receiving program when it becomes available to process it. Without using a common message transport and queueing system such as this, each application must be responsible for ensuring that the data sent is received properly.

Maintaining communications and data exchange between different types of applications can create an enormous programming burden in large enterprises. Proven middleware technologies such as IBM MQSeries provide a common programming interface (API) for which programs are written to allow them to communicate with each other across different platforms. The Transport Flow level, which includes communication protocols, message queues and message brokers, is ground zero for any EAI initiative.

Data Flow

Data flow refers to interfacing directly with the physical application data as opposed to business model level interfacing which involves integration with a public interface exposed by the application. The data flow level is one of the most important EAI infrastructure layers as key corporate and customer data is usually stored in a variety of incompatible databases across an organization. Enterprise-wide data flow is essential for maintaining a 360-degree view of customer relationships and corporate operations. The Data Flow layer is a critical prerequisite for achieving EAI success at the higher levels of information, process and business-to-business flow.

A new breed of data-centric flow tools is emerging to offer expanded EAI capabilities including real-time, bi-directional data integration for e-Business, business intelligence and customer relationship management. Obtaining data directly from a database has traditionally been achieved through standard commercial load/unload tools provided with the database or, if dealing with a proprietary storage format, by writing a program specifically for that task. But many organizations have learned the hard way that data movement and integration are much more complex than basic load/unload. The data must be cleansed and transformed into an appropriate format for access. Care must also be taken when the data is in transit between two applications so that errors or integrity problems are not introduced by the data transfer program.

The way companies think about data, and the way it is represented in databases, has changed significantly over time. Obscure naming conventions, dissimilar coding for the same item (e.g. number representation as well as character based codes) and separate architectures are all commonplace. ERP systems often use proprietary data structures that need to be cleansed and reformatted to fit conventional database architectures.

Rows and columns may have to be split or merged depending on the database format. For example, an ERP system may require that “Zip Code” and “State” be part of the same column while your company’s data structure in Oracle may have the two columns separated. Similarly, your company may have “Product Type” and “Model Number” in an inventory database as one column, whereas the ERP system requires them to be split. Data flow tools require the capability to conduct individual tasks such as translating values, deriving new calculated fields, joining tables at the publisher or the subscriber, converting date fields, and reformatting field sizes, table names and data types. Visualization (aggregate, consolidate, summarize data values), and preparation for loading multidimensional databases may also be required.

A wide range of product categories have emerged that support the movement of data between applications. According to industry analysts, there are six types of products that do this: file transfer, copy management, data propagation, schema-specific data synchronization, database replication and extraction/transformation. To reduce this to a more manageable size, the concern need only be with products that are capable of getting data directly into and/or out of an application’s data store and can also cope with changing the format of the source data to fit that of the target. This only leaves data replication/extraction/transformation products. Within this category there have traditionally been two different types of tools: code generators and transformation engines.

Code generators assist with the manual coding of programs for extracting data from an application and transforming it for loading into another application. Though the tool itself may be independent of the source or target system, the resulting program is not. Consequently, any system that is interfaced to more than one system will have considerable extra processing to perform. Another issue is that the generated program rarely has all the desired data movement functionality. So, modifications to the generated code are either performed by the user if the user can work out what the code is doing—or

through the consulting arm of the vendor. Also note that the language used for the generated program may differ from system to system – for example COBOL on MVS, but C++ on UNIX. Development staff needs to be fluent in the language of the generated program and modifying an application can also require major modifications to existing interfaces.

Transformation engine/hub tools, like code generators, use application metadata to create extract, transform and load programs. The main difference is that all the code is executed at a central location, independent from the source and target. The transformation engine/hub works by getting the source data, then moving it to a separate place (usually a different machine) where the transformation can take place. Some tools do the transformation in memory, and consequently, these tools are not scalable. For very large volumes, some tools have a transient data store option. The transient store is used for excess data that cannot be processed in memory alone, or when multiple data feeds are required. Companies may also choose among peer-to-peer solutions that perform simple data transformations on-the-fly without requiring data staging on a central hub.

Until recently, data-level tools were primarily bulk copy utilities designed for periodic data warehouse replenishment but pressed into service as data integration solutions. Recently, these traditional solutions have been joined by a new class of data flow or change data capture, transform and flow (CTF) tools that provide enhanced EAI capabilities. These tools are designed to address the fact that data-level integration has progressed far beyond the relatively simple need for batch updates. New real-time and near real-time data flow products with broad cross-platform capabilities offer features designed specifically for event-based and scheduled data integration, leading to greater information interoperability across multiple applications.

Information Flow

This flow level includes pre-built application adapters that allow core business applications such as ERP, Supply Chain Management (SCM) and legacy systems to be seamlessly integrated with applications that provide outward-facing connectivity to suppliers, partners and customers, such as CRM, e-mail systems and web-based applications. Pre-built interfaces and connectors are many times faster than custom coding and can allow EAI initiates to be implemented in record time. The goal is to rapidly integrate application and data components to form a fully functional, powerful, data-rich infrastructure and a seamless corporate IT environment.

Information flow is usually accommodated through hub-and-spoke EAI solutions that enable the movement, transformation and integration of data between heterogeneous environments from a central point of control. The main benefit of the hub-and-spoke approach is that companies can build new applications that reuse components of existing applications while providing a structured process to deliver data across mixed-system environments. This approach facilitates both bulk transfers and near real-time data flows via straight-through processing (STP), as well as message queues, one of the fastest growing segments of the EAI market. Consider a solution that supports IBM MQSeries, the dominant messaging standard in the industry today.

Hub-and-spoke information flow solutions deliver unmatched performance for customers with hundreds of application interfaces, each of which demands the highest possible throughput. The hub-and-spoke EAI architecture also allows very complex data transformations to be performed at the hub. These intelligent EAI solutions should also incorporate support for XML and HTTP to ensure that e-Business transactions can be readily integrated with the company's other core systems.

Process Flow

The Process Flow level is sometimes seen as the holy grail of application integration. It is also the most difficult level of integration to attain. Process flow integration requires integration at all of the flow layers in the hierarchy already discussed: transport flow, data flow and information flow. Process integration encompasses a range of solutions that must work together to allow business managers, business analysts and decision makers to directly define, model, manage and change business processes through a graphical interface. Business managers rather than IT managers model the process flow and thoroughly test the model before an integration solution is ever built to accommodate it.

Decision matrixes provide business managers with even greater control over process flow by providing graphical tools for the prioritization and evaluation of various process flow models. Business managers can thoroughly test process flow models and rate the relative impact or importance of various changes and alternatives. Once the integration solution is actually implemented, process flow also involves such facilities as Alarms and Alerts which allow business managers to monitor the process flow across the enterprise and respond rapidly should any problems occur.

Alerts and Alarms provide automated error notification and handling procedures that can notify administrators automatically via e-mail, wireless devices or messaging to an operator's console should selected issues arise. Support for Alarms and Alerts enables users to set up automatic actions based on conditional events. Support for Alarms and Alerts can save your organization time and resources while improving quality of service and overall data resiliency. EAI tools that include this functionality enable you to troubleshoot process flow issues immediately before they can impact business operations.

Business-to-Business (B2B) Flow

Beyond intercompany process flow integration, the real holy grail of EAI may well be business-to-business integration. While EAI initiatives have traditionally focused on back-office functions, new competitive pressures have compelled organizations to deliver the benefits of integration not only to the front-office but far beyond. Developing an EAI infrastructure that extends beyond the walls of the organization to encompass relationships between decision makers, employees, customers, suppliers and partners is essential for organizations to achieve 21st century competitiveness.

Every business-to-consumer transaction generates multiple B2B transactions—credit checks, automated billing, purchase orders, stock updates and shipping. Over the next ten years, analysts expect to see trillions of dollars in economic transactions between organizations move onto the Internet—purchase orders, invoices, cheques, legal documents and contracts. The next phase of the e-Business revolution involves achieving the integration required to let business associates transact directly with your corporate systems. This is accomplished by linking transaction systems and databases that drive the business—inventory, accounting, purchasing—with web sites and e-Business applications.

Effective B2B flow solutions use the Internet to judiciously share confidential information with business associates and eliminate the need for paper trails and redundant data entry. If your business is currently sharing information with your business partners and employees using reports, fax, FTP, EDI, mail attachments or any other means, you are missing out on opportunities to improve business processes and enhance the overall customer experience.

B2B flow tools allow organizations to filter and publish corporate data securely over intranets, extranets or the Internet so that authorized business partners with web access can subscribe to the data. These solutions must have the ability to capture information from a wide variety of corporate data sources.

Once the data is captured, the B2B flow tool must allow business partners to transform and integrate relevant information into business/desktop applications or databases on their own systems.

The solution needs to leverage the Internet and an organization's current infrastructure to provide integration and flow of data including product catalogues, price lists, production schedules, and sales forecasts to diverse systems. These diverse systems may be ERP or CRM applications supported by a partner organization or can be electronic market places for different market segments.

Evaluation criteria include near real-time integration, out-of-the-box integration with leading database technologies and desktop applications such as Microsoft Outlook and Excel, and support for open standards such as XML or Java. Open solutions that leverage existing infrastructures eliminate the need for private infrastructures, process re-engineering and costly implementations. This openness and flexibility helps businesses reduce B2B transaction costs, shorten cycle times and enhance customer experiences.

Other Considerations—Real-time Data Integration and Resiliency

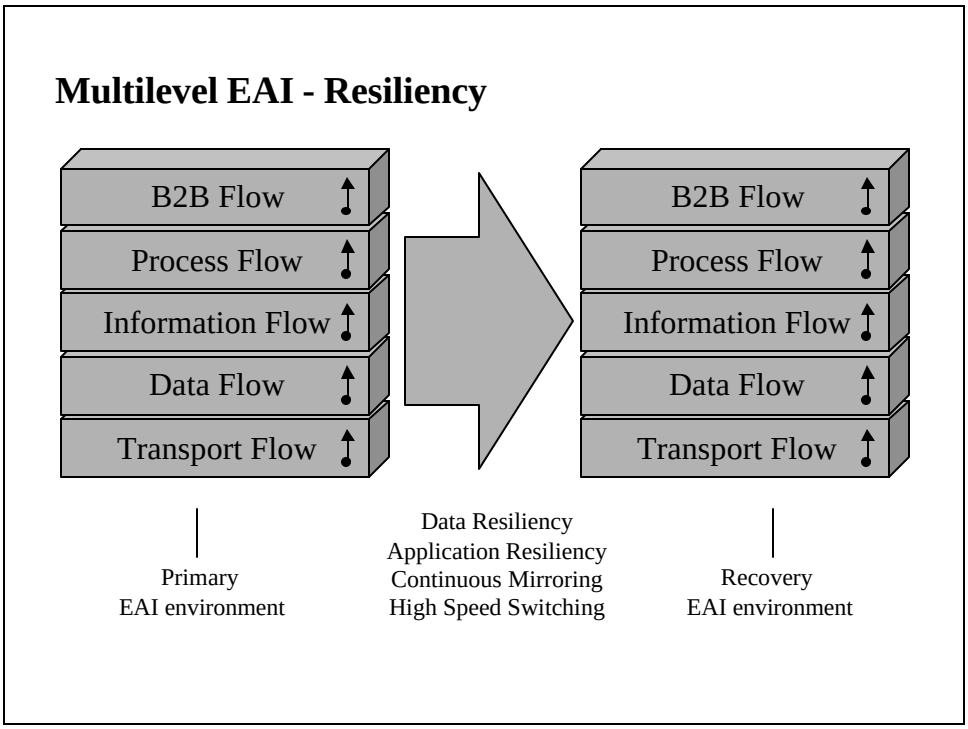
Real-time Data Drives All Business

Gaining access to the right data is only half of the challenge. When serving customers via the Internet or traditional means, the speed at which people and applications can access the right data is equally important. A critical question that must be answered before application interfaces are built is: how fresh or current must the data be when it arrives at its destination? The need to send data as quickly as possible is fundamental to e-Business. In fact, no sooner has the data been processed at the source than it must be sent on to other applications that need to process it. Moreover, the real-time e-Business model is transforming the expectations of end-users and decision makers in other areas of the business as well such as business intelligence and CRM.

Organizations have traditionally supported customer service operations with a mixture of service levels, ranging from the slowest (batch) to the fastest (real-time). These have typically been defined by the importance the organization gives to the information. For example, a retail bank needs to have the ability to send real-time information to its ATM networks to prevent fraud. However, many back-office account reconciliation functions only require overnight batch processing. While no one would expect every business process to change to support the new generation of real-time systems, the general movement of IT systems is toward the real-time or near real-time model. There is a clear need for on-line channels to perform business functions in real-time, based around continuous interfacing with legacy and batch-oriented back-office systems.

Resiliency

Recently several high profile companies in the e-Business space have been plagued by repeated system downtime. The Internet has given rise to a new breed of customer that demands 24/7 availability. If your web store isn't open, customers will simply point their browser elsewhere and go there with the click of a mouse. A corporate web site is now the business equivalent of a company's front lobby, on-line store front and help desk all rolled into one. 24/7 requirements for customer experiences have also spilled over into the bricks-and-mortar world where customers now expect instantaneous and uninterrupted service from virtually all customer touch-points.



Caption: In addition to facilitating integration at all flow levels, companies must also ensure the availability and resiliency of critical data and applications.

A system outage or database crash can also seriously undermine your ability to collaborate effectively with partners and suppliers. A long outage can cause your supply chain processes to grind to a halt. Your business data and applications are precious and making sure they are available to customers, partners and employees around the clock is critical. It is absolutely imperative for companies to build a highly resilient distributed EAI infrastructure that allows organizations to safeguard their business data and application assets.

High availability solutions are available that provide the ability to continuously mirror critical data and applications and automatically switch users from primary to recovery systems in the event of an outage. This functionality not only enables customer-facing systems like CRM and e-Business to be thoroughly backed up and resilient 24/7, but also ensures that application integration environments which connect internal systems with business partners and suppliers remain continuously intact. The overall goal is to create and maintain a zero latency enterprise application integration infrastructure that is virtually impervious to downtime interruptions.

Summing Up

An effective multilevel EAI strategy empowers organizations to achieve new operational efficiencies while enhancing customer service levels and maximizing revenue opportunities. With the introduction of new tools, methodologies and approaches, enterprise application integration is evolving to meet the changing needs of business. Now more than ever, companies have at their disposal a number of viable alternatives for large scale and complex EAI initiatives that involve messaging and business rules integration. The challenge for software decision makers is that there is often considerable crossover and overlap between these technologies. As the capabilities of EAI tools evolve and as the boundary between real-time data flow solutions and near real-time messaging tools is blurred, there are many instances where either may be sufficient for a particular interface.

The recommendation to IT managers is that these new EAI tools are worth evaluating in relation to the multilevel approach outlined above—provided they support open standards like MQSeries, XML and Java. Thorough testing from proof-of-concept onwards will determine whether they are a good fit with an organization's application network.

About DataMirror Corporation

DataMirror (Nasdaq: DMCX; TSE: DMC) delivers solutions that let customers integrate their data across their enterprises. DataMirror's comprehensive family of products includes advanced real-time capture, transform and flow (CTF) technology that gives customers the instant access, integration and availability they demand today across all computers in their business.

Over 1,300 companies use DataMirror to integrate their data. Real-time data drives all business. DataMirror is headquartered in Toronto, Canada, and has offices worldwide. DataMirror has been ranked in the Deloitte and Touche Fast 500 as one of the fastest growing technology companies in North America.

DOWNLOAD

EVALUATE

IMPLEMENT

DataMirror software is available on-line at datamirror.com or call 1-800-362-5955

DataMirror®