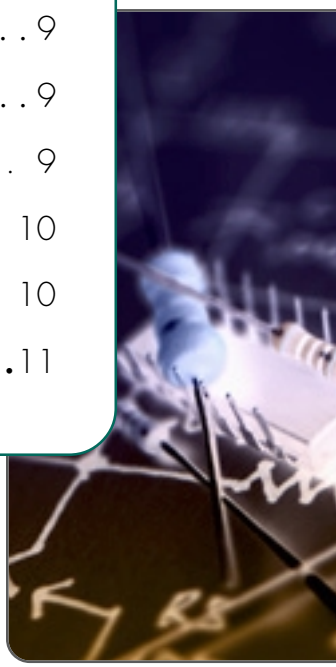
Data Routing In the Age of Information

December 2001

The HTRC Group, LLC
P.O. Box 2087
San Andreas, CA 95249
www.htrcgroup.com

Data Routing Whitepaper



Introduction - Current Landscape

Growth of information and Web applications

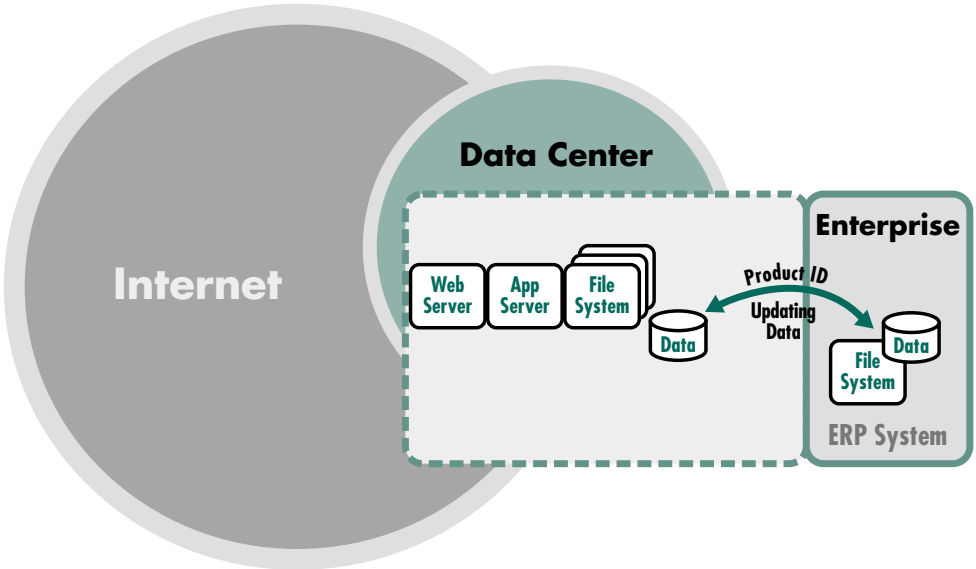
The growth of information on the Web is staggering. Enterprises across all industries are rapidly deploying Web-based applications as a means to more effectively communicate and distribute information to their customers, partners, and employees.

Today's Web-based applications are extremely varied. Application types often range from Enterprise Information Portals, to B2B Exchanges, to Extranets, to E-Commerce sites, to Content/Interactive Communities. Moreover, these Web-based applications have grown dramatically over the past few years—from largely static Web sites to complex distributed applications that are deployed across multiple data centers, and offer highly personalized and dynamically driven content.

Increased data sharing between systems

Web-based applications often use data from several existing enterprise applications within and outside the corporation. Nearly all of these applications require new and dedicated systems (e.g. databases, application servers, Web servers, etc.) that must be uniquely deployed, managed, and maintained. As a result, new processes are required to transfer data from these existing enterprise systems to the new Web-application infrastructure.

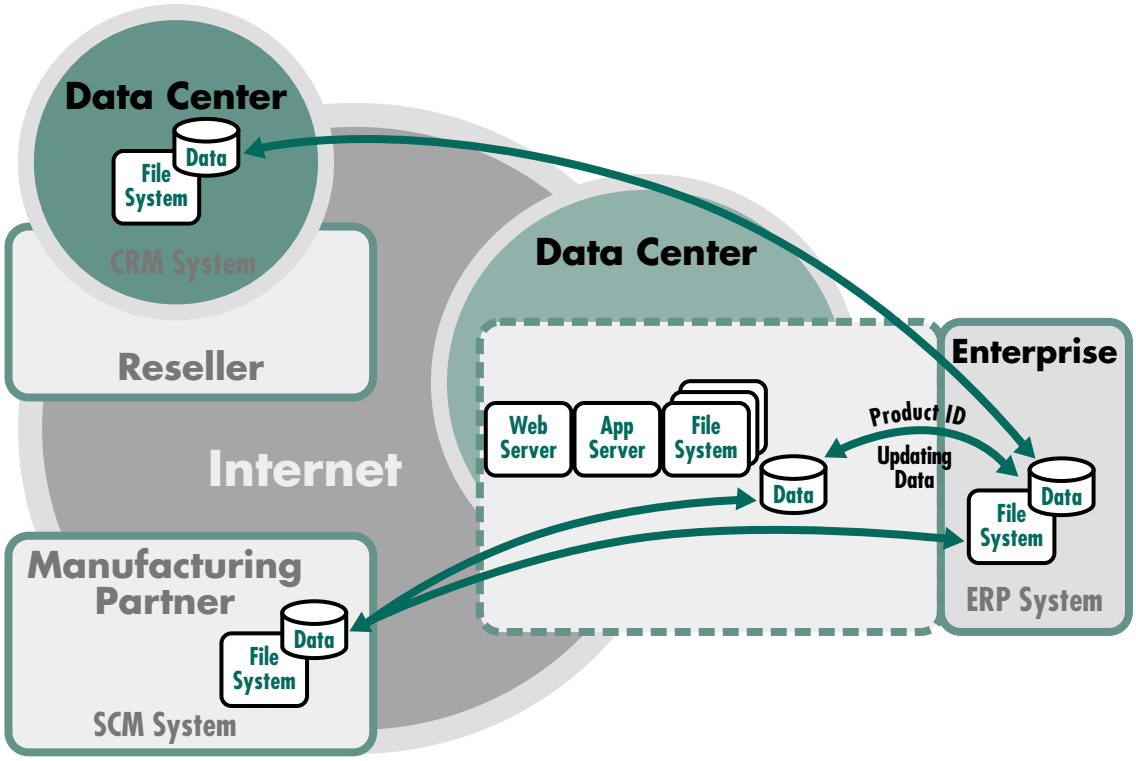
For example, a Web-based application that delivers product descriptions will likely use the same information as an enterprise sales application. Changes and updates to the enterprise data system will need to be reflected in the Web-based application data system. Managing the process of updating Web-based applications from a distributed set of data sources (especially as the number of both Web-applications and data sources continues to grow) can be extremely inefficient and costly.



The practice of deploying Web-based applications on separate systems and infrastructure, however, is very likely to continue. One of the primary reasons for this trend is the heavy reliance Web-applications have on databases. Separate databases are developed for new Web-based applications for several reasons:

- Separate groups developed the application, and used different database solutions.
- New business process requirements with specific data requirements are each implemented with their own software and database solution.
- Larger companies tend to have multiple geographic locations, and many of them run corporate-wide enterprise applications by aggregating data from a variety of local data stores.

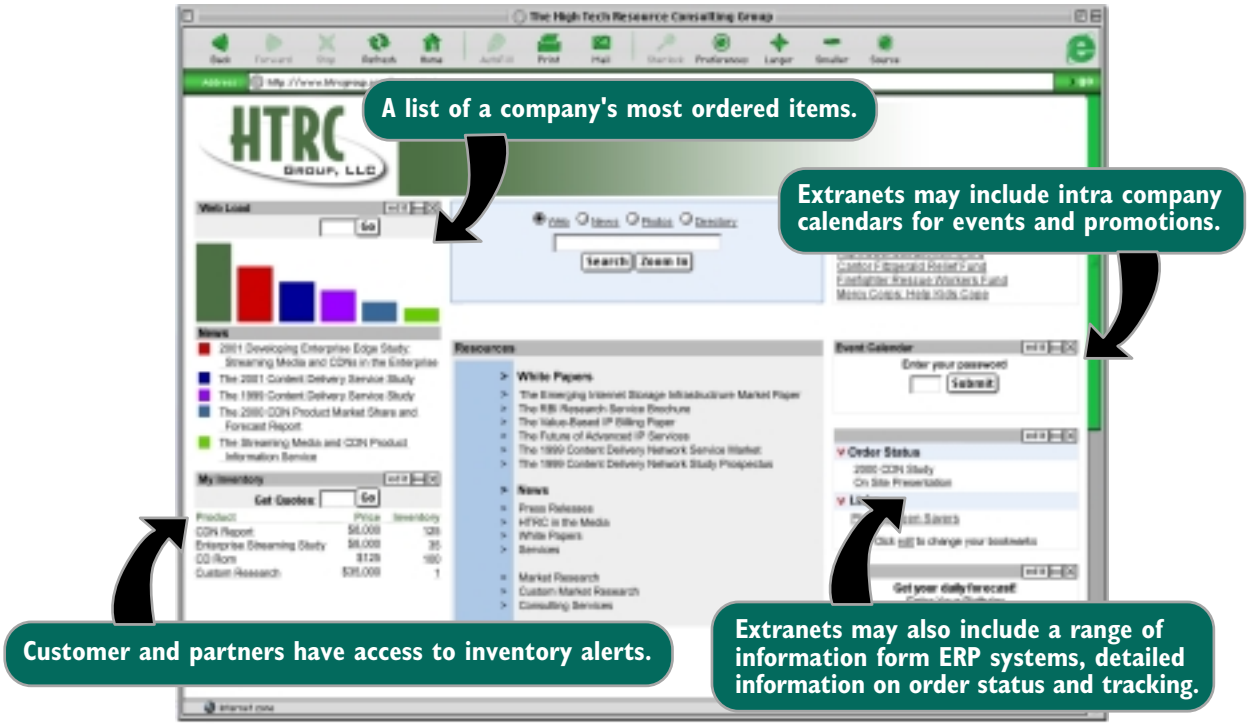
The examples mentioned above have only discussed the process of new Web-based applications using data from internal enterprise systems. However, today's Web-based applications, commonly require data that is aggregated not only from internal systems, but also from many external organizations and systems.



Extranets provide businesses with a way to effectively communicate with customers, suppliers, and partners. For example, a corporate Extranet facilitates critical communications with customers, enabling customers to check on inventory information through their supplier's Extranet. Web services on an Extranet likely include the integration of many applications that reside in the private enterprise Intranet, the Extranet, and the public Internet. These applica-

tions may include inventory alerts, order status, help request status, lists of the most ordered items, partner company news, and a product search area. Our example below shows how some of these applications would be integrated into an Extranet.

In fact, the overall vision of collaborative computing is based on the premise that applications should be able to share information freely across all boundaries whether organizational, geographical, or technical. Over time, as corporations continue to deploy Web-based applications that rely on a composite of data aggregated from both internal and external systems, the complexity and costs associated with the sharing of data between these coordinating applications is destined to increase exponentially.



Rapid-Build Out of Web-based Applications

Unfortunately, the vast majority of today's Web-based applications were designed, built, and deployed at break-neck speed. Fierce competition, coupled with the Internet economy's incredible rate of change, forced companies to make very quick technology decisions. As enterprises focused on simply getting the application up and running, ad-hoc processes for moving data in and out of these new systems were developed quickly.

Looking back, there is no wonder why many companies, both dot.coms and brick and mortar, have engaged in n aggressive build-outs of stove-pipe applications, each with a separate function. However, this pressure to remain competitive has engulfed many companies and consumed those that made fatal technology decisions. Homegrown processes and tools are now proving to be extremely rigid and brittle.

Nevertheless, the first wave of innovation gave many a glimpse of how powerful the IP connected world could be—a vision of a future connected world and the opportunity it presents.

Emerging Technology Trends

Fortunately, several important technologies and standards have emerged to help address this new set of challenges. These technologies build on the hard lessons learned over the past several years and have been specifically designed to facilitate new levels of interoperability, and to catalyze the next wave of Internet computing.

XML

According to The World Wide Web Consortium (W3C), "The Extensible Markup Language (XML) is the universal format for structured documents and data on the Web." According to The HTRC Group's "The 2000 Content Delivery Service Study," the fastest growing content type is content created with XML with 27% this year growing to 67% in 2001. XML enables the flexible transmission and interpretation of data between applications. As enterprises increase their exchange of data/information to create new Web-based applications and services, XML offers a new level of data interoperability.

Web Services

Web Services represents another significant class of technologies that have emerged to provide greater means to integrate disparate applications and services together. Web Services provide a common framework for finding, invoking, and integrating distributed components/services over the Internet. What makes Web Services particularly unique is that its framework utilizes the Internet's current open data standards and protocols, such as XML and HTTP, unlike earlier component models such as DCOM and CORBA that relied on proprietary object models and protocols. With Web Services, components that are implemented in multitude of different languages and platforms, and that reside in different locations in the network, can now be loosely coupled together to form distributed applications. To this end, Web Services provide a standards-based mechanism for applications to programmatically request information from other applications over the Internet.

XML and Web Services are focused on enabling the next wave of Internet computing, often referred to as "the semantic Web", where applications are able to easily share information and work together. These technologies, however, are not solutions in and of themselves.

Moving From Packet Routing to Data Routing

The need to automate the process of moving data between multiple sources and destinations, and to support a growing number of protocols and standards that are fundamental to this process, is very analogous to the industry evolution of packet-level routing/networking.

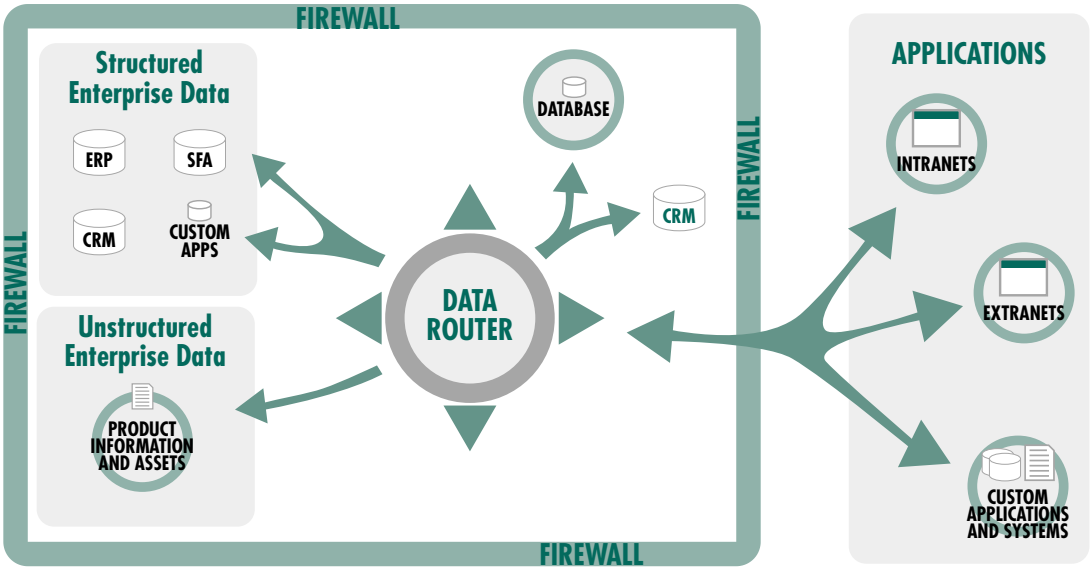
Packet routing has given us a world where any IP-enabled computing device can become part of the network of networks. However, packet routing deals primarily with moving bits, or "1s and 0s", around the network. The next step is to enable the movement of data with meaning. For example, common items such as <shipping order>, <product id>, or <billing address> should move freely and securely between applications and systems. By providing a

foundation for the movement of semantic information, data routing has the potential to become the next major wave in computing. Data routing will enable new levels of data interoperability and business process automation.

Data Routing

Data Routing represents a new class of solutions that leverage XML and Web Services to help automate the process of sharing data/information between systems. While XML and Web Services provide the standards and protocols necessary to more easily exchange data, data routers will ensure that the flow of data between systems is handled in a reliable, secure, timely, and scalable fashion.

Data routers are systems that automate delivery of data from multiple sources to multiple destinations. They facilitate the movement of multiple data types whether they reside in databases, file systems, or enterprise applications. Data routers allow companies to index relevant data, facilitate data transfer over open standards protocols, and programmatically access the router via Web Services.



Core Functionality

Provide access to data sources

Data routers have the ability to abstract and index the data to which they provide access. Data sources such as file systems and databases are abstracted from requesters so that requesters need only submit the name of a resource to receive the appropriate data. Without this capability, users would be forced to submit entire complex queries or specify the exact location and filename on a file system. Separating the actual query from the name of the resulting dataset enhances usability of data by eliminating the complexity of requesting data.

Configure policies for moving data

Another core function of data routers is the ability to configure and manage the movement of data. In addition to enabling the specification of data sources and targets, data routers also allow users to build complex routes that can include multiple transforms, sources, and destinations. Routes can be triggered by specific events, calendar-based, or one-off requests for data. Data routers provide additional controls for encryption, authentication, caching, and other options. Moreover, data routers provide queues, which provide for policy-based traffic prioritization and active queue management.

Provide programmable and Standards-based Interface

Data routers provide a wide variety of means by which to request and receive data. Requests for data that are issued via protocols such as SOAP, XML-RPC, and JMS, are interpreted via standards-based interfaces. Because data routers have Web Services interfaces, requesting applications can communicate with the router over HTTP (port 80) and avoid any conflicts with firewalls. Moreover, data routers' scriptable interfaces provide for extensive flexibility and automation.

Move data securely

Because of their strategic location and function, data routers provide a level of access control policies in addition to existing security policies. Privileges for users and applications are checked with the data router to ensure only authorized usage. Additionally data can be routed over open standards based security protocols. Secure data routing enables multiple Intranet, Extranet, and Internet applications to securely use multiple data sources throughout the public and private enterprise network.

Case Study

A telecommunications reseller sells usage of networks run by several large telecommunications providers. The telecommunications reseller has built an internal database for customer management. They wish to aggregate the summary billing information received from the telecommunications providers (received in XML) with the customer database in order to provide their customer representatives with quick access to the information through a Web UI. Furthermore, the company wants to make the data accessible via their online service portal.

The data router is configured to retrieve XML data from each of the providers, apply transformations to break out billing information by customer, and then move the resulting data into both the CRM database and the customer service portal database. The data router's support for authentication and security protocols makes movement of XML data from providers to the target databases secure. As new billing information becomes available each month, the router automatically updates the database.

The reseller also wants to offer partners or customers direct interfaces to billing and reporting data. The data router allows those organizations to build their own custom applications that use Web Services protocols to access information. As a result of native Web Services support, the reseller can easily integrate new partners. The data routing solution provides the reseller with a secure, scalable mechanism to share data with its partners.

Benefits

Business Process Optimization

Synchronized policy and data flow

Data routers help align technology with business processes, marrying business policies with data flow and applications. These access and routing control policies enable multiple Intranet, Extranet, and Internet applications to optimally move data where and when it is needed throughout the public and private enterprise network. Data routers enable the automation of tasks that rely on dynamic business conditions, thereby allowing organizations to change static legacy systems into intelligent, dynamic systems.

Extended data reach

By making data available through open standards-based protocols, data routers make enterprises' data easily accessible to far more applications, whether they are inside or outside the firewall. Companies can use data routers to effectively leverage their business information with potentially all the links in their respective value chains. Extending the reach of enterprise data saves time and money, and opens up new possibilities for the wealth of data accumulated by organizations.

Automated data flow

Automation simplifies and accelerates business processes. Managers can use data routers to configure routing policies to specify when and how data will flow. Routing policies can be as simple as moving a file every night at 11 pm, or as sophisticated as making the results of a complex query available to external business partners only when a specific event occurs. By automatically completing data tasks based on set schedules and events, data routers reduce the number of errors associated with unnecessary human intervention in manual tasks. This quickens the pace of business execution, and saves money.

Infrastructure Optimization

Network

The data router reduces the amount of unnecessary point-to-point traffic between disparate data sources and targets. By registering data in the data router at the required atomicity, data routes can be configured to move only the specific required subset of data, eliminating redundant data transmission and thereby reducing bandwidth costs. Another function of a data router is to increase application performance through database and file system caching. By caching frequently used data queries, the data router can improve database and file system performance across multiple systems.

Storage

Data storage needs are constantly increasing. Respondents from the "Developing Enterprise Edge" study average 12,487 GB in 2001, increasing to 20,397 in 2002. In the study, storage requirements increase 7,910 GB from this year to the next. The implementation of a data router solution can reduce storage growth requirements to more manageable and cost effective levels by optimally utilizing existing file servers and disk storage. Data routers help optimize storage by accessing data directly from multiple sources, reducing the need for redundant data storage and replication.

Delivering ROI through Cost Savings

The need to deploy applications cost effectively has created a need for data routers. Data routers can enable businesses to cut costs by leveraging existing business data and feeding new applications. Furthermore, they can eliminate the need to deploy entirely new databases and repositories. More significant than the cost of deploying additional data stores, is the cost of managing the redundant data stores. Additionally, data routers reduce the need to write custom code to populate applications and data stores, further decreasing costs. As companies rush to deploy new Web applications, they are coming to grips with the increasing costs and complexity associated with feeding these new applications with data. Technology solutions adopted in order to support these Web applications must provide clear value and a solid return on investment. The data router provides a cost-effective means to access, move and track data throughout growing enterprises.

Conclusion

Over the last few years the proliferation of Web-applications to support networked business processes has led to the vast need for integration between Web applications and data systems. Data integration solutions evolved from one-off in-house implementations to simple ETL tools, replication, federated database solutions, and highly complex EAI systems.

In parallel, increased standardization of business data representation is fuelling proliferation of inter-operable Web-applications. Today, XML is rapidly becoming the lingua franca for business data, and penetration of data semantic standards such as RosettaNet and ebXML is increasing quickly. With the standardization of semantic data representation it becomes possible to also standardize on data services. Web Services, based on open-standards like WSDL and UDDI for service definition and discovery, are developing as the dominant trend in enterprise computing in this decade.

The increasing need for interoperable Web applications coupled with the growing adoption of open data standards has created a need for data routing. Data routing provides the opportunity to create a standardized set of solutions that use Web Services technologies to reliably and securely share data between systems. Data routing is fundamental to supporting business processes within an organization - whether it is moving unstructured document files between file systems in order to support interdepartmental processes, populating new Web databases to support a new portal initiative or providing new Web Services to offer request-based transaction log access to trading partners.

Data routing will be as fundamental to enterprise business processes as packet routing is to networking. Data routers are the equivalent of packet routers for semantic data. Data routers enable the secure, scalable and optimized routing of semantic data at the right place and time, not only for immediate cost optimization but also for the long term robust build-out of new interoperable business processes in today's and tomorrow's real-time enterprise.