Inside Gartner Group
19 August 1998

Outflank the Competition by Deploying IT to Build a ’Zero-Latency Enterprise’

B. Enslow, R. Schulte

Zero-latency information strategies drive instantaneous awareness of supply chain events across an enterprise, which can enable the enterprise to outflank its competition.

Enterprises that act quickly on new information have a competitive advantage, and "zero latency" strategies aim to push that advantage to its ultimate limits. Zero-latency information strategies drive instantaneous awareness of supply chain events across an enterprise, which can enable the enterprise to outflank its competition.

Enterprises with quick reaction times have competitive advantages that yield many benefits, chief of which are:

- Selling more products or services
- Providing better customer service
- Capitalizing on new business opportunities faster than competitors

There are many ways to speed up the flow of information and business processes to achieve some degree of those benefits. Zero-latency-enterprise strategies, however, take the goal of information timeliness to its ultimate limits: The goal is instantaneous awareness and appropriate response to events across an entire enterprise or beyond.

However, enterprises must be aware that zero-latency-enterprise strategies are appropriate for only certain business processes and require the adoption of emerging messaging and integration tools - no zero-latency system is available "off the shelf."

The Benefits of Zero Latency

In technical terms, "latency" is the time a system takes to respond to an input. If one views the modern enterprise as a kind of complex system, a zero-latency-enterprise strategy implies that all parts of the enterprise can respond to events as soon as they become known to any part of the enterprise. The various divisions, departments and even groups within external business partners are treated as cooperating "subsystems," regardless of where they are located. As soon as new information is captured by any application system in any workgroup, it is made available to all other interested parties.

In some cases, the interested party may be an application system - e.g., on a mainframe or some other server - that will process a transaction, update a database, and perhaps send a follow-up transaction elsewhere. In other cases, the interested party may be a person using a browser or a desktop client/server application.

Sharing information in real time among disparate applications and people provides many supply chain management benefits. For example:

- A transportation carrier can tell a waiting customer whether a shipment has cleared customs.

- An inventory system in Minneapolis and a sales reporting system in Chicago can finally agree on the number of items stored in a warehouse in Buffalo.

When the "right hand" knows what the "left hand" is doing at every moment, the possible efficiencies and opportunities for enhanced customer service and revenue opportunities are enormous.

Such benefits are definitely achievable, but real-world implementations are necessarily limited in scope. Leading-edge enterprises - such as 3Com, Bechtel, Delta Airlines and Federal Express - have already achieved substantial benefits by applying the concepts across a range of their applications. Each enterprise has re-engineered business processes, tailored certain application systems for real-time operation, and implemented a broad messaging infrastructure.

Of course, it would be impractical and pointless to redesign every function in a large enterprise. Not all business problems can be helped by more-current information. Still, it is extremely valuable to accelerate the information flow for selected aspects of the enterprise using the principles that underlie a zero-latency strategy.

The Five Components of a Zero-Latency Strategy

A successful zero-latency strategy requires five key components:

1. A network and software infrastructure that is capable of quickly exchanging information across technical and organizational boundaries

2. End-user interface tools - e.g., browsers - and other application programs capable of sending and receiving information through the infrastructure in a timely fashion, and applications or human processes capable of dynamically analyzing and acting on the information

3. A business strategy that leverages velocity as a competitive weapon (see the sidebar, "Tools for Increasing Velocity")

- Data and planning velocities in many enterprises fail to support the pace that the business needs to be competitive. The result can be excessive order fulfillment, product conversion, order to cash, and product commercialization cycles.

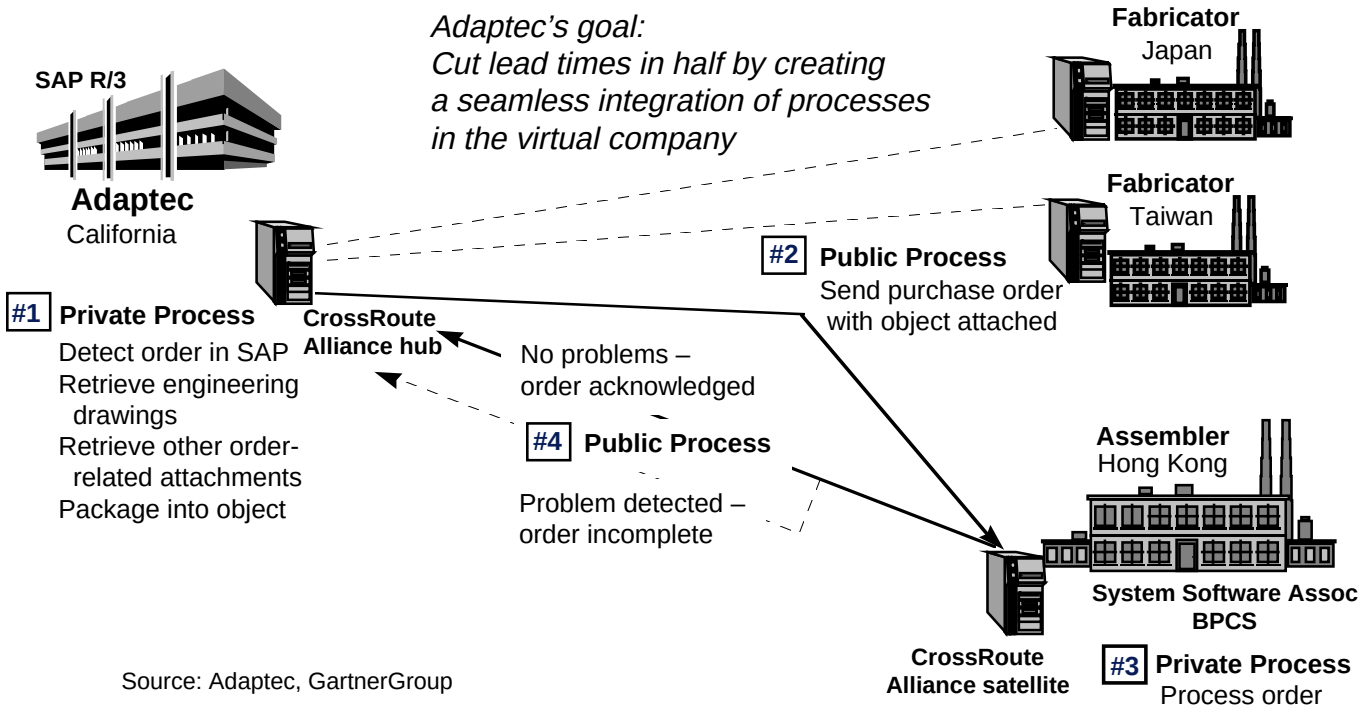
- Enterprises striving for market improvement or dominance use velocity management as a key tactic.

- For example, California-based Adaptec implemented Internet-based processware from CrossRoute Software with Adaptec’s contract computer chip manufacturers in Asia (see the sidebar, "Processware: Instant Application Integration"). The implementation has enabled Adaptec to achieve real-time, application-to-application communication and cross-application process flow, which radically improved data velocity compared to the old process of overnight

batch electronic data interchange (see Figure 1). The benefits include a 33-percent reduction in order fulfillment cycle time and inventory savings of $10 million in six months.

Figure 1

Adaptec Increased Data Velocity With Processware From CrossRoute Software



- Dell Computer, which measures inventory in hours, not days, is so focused on velocity as a competitive weapon that it has coined an internal word for it: "Dellocity."

4. A set of business policies, processes and even product offerings engineered to implement the business strategy

- Zero latency can drastically affect the way tasks are done and how goods and services are delivered.

- For example, zero latency may change product allocation and packaging, alter service procedures or sales methods, change sign-off procedures, eliminate reports or forms (e.g., eliminate purchase orders to suppliers), or add new procedures and forms (including Web pages).

5. An enterprise that can implement the new processes

- The structure and function of workgroups and departments may have to change.

- Zero latency potentially affects the work of all job categories, from the worker on the loading dock all the way up to senior executives.

- Contact with IT resources can change considerably.

The last three components are business issues that involve considerations beyond the scope of this article. Consequently, only the first two components are addressed below.

Zero Latency Means "Fast Enough"

Zero latency does not imply any specific technical architecture. GartnerGroup considers a zero-latency strategy to be any strategy that exploits the immediate exchange of information across geographical, technical and organizational boundaries to achieve business benefit. Technical boundaries exist between, for example, different operating systems, different database management systems and different programming languages.

How fast is immediate? Immediate implies being fast enough to bring all the business benefits that could potentially be achieved by simultaneous knowledge. Latency cannot really be zero in any system, of course, because computers need time to process programs and every transmission link involves various delay times across the network links. For most business tasks, all the tangible benefits of simultaneous knowledge can be achieved by an infrastructure that can exchange information as fast as the response time of a typical online application, which is generally 0.4 to 2 seconds. Moreover, humans are often an integral link in the system - inputting, analyzing, or reacting to the information - and humans certainly cannot work faster than computers.

Zero latency should not be deployed blindly. Enterprises with business problems that can be fully satisfied using information that is hours old or even days old do not require a zero-latency strategy. For such enterprises, an enterprisewide integration infrastructure can use batch file transfer or even manual interchange steps. Consequently, enterprises must look at the time constants of their businesses to identify where zero latency makes sense, and ensure that data velocity is synchronized with planning and execution velocities.

Designing for Zero Latency

The purpose of a zero-latency communication and software infrastructure is to reduce information float. Information float is the time between when data is captured in one place and when it becomes available and usable elsewhere.

Two basic design patterns are commonly used for zero latency:

- Event-based messaging is becoming popular in supply chain management vendor solutions because it minimizes overhead in many situations and theoretically has the lowest latency. A business event refers to any meaningful change in data or conditions, such as a new order submittal, a change of address, an airplane landing or a package delivery. An event-driven (or "push"-based) system is one in which tasks are triggered when a program or a person receives notification of the change. The sender determines when the information transfer will take place.

- A zero-latency operational data store is a shared database that receives updated information as it is generated by applications running anywhere in the network. When a person or application program needs the latest information on a subject, an inquiry is sent to the database, which returns the answer. This is a request/reply or "pull" communication pattern in which the recipient determines when the information transfer will take place.

In practice, zero-latency solutions often mix the two design patterns. For example, an operational data store may be updated by event-based messaging, but inquiries into the data store may be request/reply. The details of the design are irrelevant. As long as the design enables the immediate exchange of information across boundaries, it is considered a zero-latency solution.

Zero latency often uses "publish and subscribe," a particular form of event-based communication in which information sources publish (send) information to a somewhat-intelligent infrastructure; information consumers subscribe by specifying the kind of information they want to receive. The publish-and-subscribe pattern is scalable, flexible and easy to modify in the face of constant changes in the application systems and user business requirements.

Achieving Zero-Latency Across Heterogeneous Systems

Most zero-latency strategies involve some form of messaging middleware. A zero-latency strategy that spans multiple departments or trading partners will inevitably encounter such a diversity of application systems that the infrastructure must also be a message broker.

A message broker is a logical hub that acts as an intelligent intermediary (broker) between application systems. It enables program-to-program communication between disparate applications and offloads logic from application programs by providing one or more of the following services:

- Transformation. Transforms messages from the incoming message format to different output formats

- Message Warehouse. Temporarily stores messages to be analyzed or retransmitted at a later time

- Flow Control (workflow). Organizes multistep business procedures

- Message Dictionary. Holds metadata description of message formats

- Administration and Monitoring. Manages the broker configuration

- Adapters. Provides tools for connecting to participating applications

Not all message brokers are built with the goal of zero latency; many are used, in part, for slower forms of application integration involving file transfer. Similarly, not all messaging solutions are brokers; the majority of messaging systems are used within the bounds of a single distributed application.

Zero latency can be implemented using a variety of technologies. The Internet often plays a role in transporting information among participants, particularly for wide-area and interenterprise solutions. A range of middleware - some that is Internet-based and some that is not - may be appropriate, as illustrated in the following examples:

- Federal Express uses Vitria Technology’s BusinessWare Communicator to track several million packages per night at its Memphis hub. By knowing the volume of packages bound for each destination minutes - or sometimes even hours - earlier than in the past, Federal Express is able to dynamically reassign workers and equipment to heavily trafficked destinations, thus speeding up turnaround times and making more-efficient use of resources.

- Delta Technology, a subsidiary of Delta Airlines, has a hybrid infrastructure that blends an Oracle data store with IBM MQSeries middleware, custom software and VIE Systems’ Copernicus transformation tool. Up-to-the-second information about every Delta flight is collected from hundreds of sources, maintained in a central operational data store, and made available to managers and other workers handling numerous operational aspects such as scheduling gates, fuel delivery, aircraft maintenance and flight crews.

- 3Com, Bechtel and the Nasdaq Stock Market use Tibco's Rendezvous as the middleware core of their infrastructures.

- Other enterprises have built their zero-latency solutions on Talarian's SmartSockets publish-and-subscribe middleware.

A communication infrastructure cannot enable zero latency by itself. Moreover, senders and receivers (programs or people) must be established to deal with information immediately. This does not imply that the participating applications are new, although some may be. Legacy and purchased applications can be modified to participate in zero latency. Alternatively, a software adapter or "wrapper" can be inserted to mediate between the application and the infrastructure - i.e., a noninvasive connection.

Many enterprises use a combination of approaches because the majority of their applications are legacy or purchased. New applications that understand and can quickly react to variability - such as dynamic routing programs - may have to be added.

The second target type, end users, is addressed directly by the emerging push vendors - e.g., Wayfarer Communications, Wall Data, DataChannel and BackWeb Technologies.

Zero-Latency: The Latest Time-Related Strategy

The notion of zero latency is best understood in the context of other time-related strategies that have appeared during the past 20 years. Zero latency can complement other management initiatives, yet it is distinctly different from any of them.

- Just-in-time (JIT) manufacturing or retail distribution strategies drive down inventory costs by reducing inventories to a low level. The supply of input materials must never dry up, but the quantity on hand must never become too high either. JIT requires the timely exchange of information between the supplier and the consuming organization. A zero-latency infrastructure is potentially a good enabler of JIT. Build-to-order manufacturing strategies that eliminate unsold finished goods inventory also may be enabled by zero-latency information.

- Straight-through processing (STP) is a popular strategy for financial institutions and is pursued by some utility companies under the name of "flow-through processing."

- STP's fundamental premise is that a transaction - such as a payment, a trade, or a change to a residential phone service - is entered only once. Thereafter, it proceeds in an automated fashion through the rest of its life cycle, which may include dozens of steps in different application systems in various locations. The goal is to avoid rekeying information, thus reducing input errors and shortening the time to completion.

- STP is complementary to zero latency and the tools and techniques overlap. Both strategies focus on interapplication communication and usually rely on message brokers. However, STP manages the progress of a predictable, multistep business transaction through hours or days. Each STP message may be logged and tracked to ensure that transactions are not lost. Zero latency, on the other hand, is based on immediate, broad (many-to-many) information distribution, and measures to guarantee message delivery are not always necessary.

- Any traditional online transaction processing (OLTP) application could be considered a zero-latency system in the sense that it allows updates to be visible immediately to other users of that application. However, it is the ability to link disparate, geographically dispersed

application systems that distinguishes a zero-latency enterprise from a zero-latency application system.

Bottom Line

When OLTP came into broad use in the 1970s, its impact on business operations was enormous. Early adopters of IBM's CICS transaction-processing monitor and other such monitors could deliver better customer service and thus gained a huge advantage over competitors that clung too long to their batch systems.

A zero-latency-enterprise strategy has the potential to achieve a similar breakthrough in supply chain management operations by applying the notion of immediate communication among multiple, independent, distributed systems.

Tools for Increasing Velocity

Enterprises seeking to increase velocity in key areas must first measure the heartbeat of their processes and identify areas in which data, planning and execution disconnects. Appropriate applications and technologies may then be applied to fix the disconnects and increase velocity.

Below is a partial listing of such tools - all of which are deployable today. Deploying the tools without synchronizing data, planning and execution velocities will result in local optimization of activities, but will fail to achieve the global optimization required to significantly improve the velocity of key business processes.

Data Velocity

- The Internet, extranets, real-time electronic data interchange (EDI)
- Message brokers
- Processware
- Enterprise resource planning (ERP) transaction system
- Advance shipment notices

Planning Velocity

- Memory-resident advanced planning and scheduling software
- Enterprise demand planning system
- Real-time decision support software - e.g., dynamic routing
- CAPE II (concurrent art-to-product environment II)/product data management

Execution Velocity

- Electronic funds transfer

- Manufacturing execution system and shop-floor control systems
- Cross-docking and flow-through warehousing systems
- Supply-chain-aware sales configurators
- Kanban (a Japanese just-in-time process) and flow manufacturing systems

Processware: Instant Application Integration

As the hype of standardized enterprise resource planning deployments for supply chain management fades, a new class of applications has emerged to satisfy the rising challenges to integrate disparate systems thorough the supply chain. GartnerGroup classifies this new type of application as "processware." This class of software, which is promoted by a growing group of vendors and standards groups, purports to solve easily and inexpensively all the problems vexing enterprises that must integrate purchased and internal applications. That need intensifies as enterprises increasingly automate, through IT deployment, the links between departments, customers and suppliers. Moreover, as enterprises continue to outsource business processes (e.g., manufacturing, help desks, warehousing and transportation), the need to integrate potentially disparate systems increases even more.

However, the capabilities and focus of processware offerings vary. Some, such as Tibco's, provide basic technologies, while others, such as CrossWorld Software's, offer specific application-to-application mapping. Many processware vendors claim that their products permit a high and somewhat seamless level of best-of-breed integration between applications. However, enterprises should understand that each application integration approach has strengths and weaknesses, and each involves varying degrees of work to achieve an acceptable solution.

There are three groups of processware approaches:

- *Interenterprise Integration.* Includes Actra, electronic data interchange vendors, CrossRoute Software and Service Track Enterprises.
- *Intraenterprise Integration.* Includes messaging middleware (e.g., IBM's MQSeries, Tibco and Vitria Technology) and message brokers (e.g., TSI International Software, Neon Systems and Frontec AMT).
- *Industry and Vendor Integration Specifications.* Includes Object Management Group, Open Applications Group, IBM, Microsoft (Value Chain Initiative), Oracle (Network Computing Architecture) and SAP (Business Framework).

The three groups have conflicting as well as complementary goals. The industry and vendor group is trying to minimize the need for the other two groups, but this goal will not be achieved by 2003 (0.8 probability). The integration tool players have varying approaches, sometimes oversold, that will require enterprise investment and maintenance. None will emerge as the true "silver bullet" for application integration needs by 2003 (0.8 probability).

Written by Edward Younker, Research Products

For related articles published in InSide GartnerGroup This Week, see:

- "The Virtual Enterprise - The Phenomenon That IT Built," 3 June 1998
- "Dynamic Enterprises Need Flexible IT Architectures," 6 May 1998
- "The Whys and Hows of Knowledge Management," 4 February 1998

This document has been published by:

Service	**Date**	**Document #**
InSide Gartner Group	19 August 1998	IGG-081998-01