

Enterprise Application Integration and Active Data Warehousing

Stephen A. Brobst
Chief Technology Officer
NCR Teradata
17905 Via del Campo
San Diego, California 92127, USA
Stephen.brobst@ncr.com

This manuscript was originally published in:
*Proceedings of Data Warehousing 2002: From Data
Warehousing to the Corporate Knowledge Center.*
Physica-Verlag Heidelberg, November 12-13, 2002. pp. 15-23.

Abstract

When data warehouse repositories were kept for exclusive access by the privileged few in marketing, finance and strategic planning there was little need to be concerned with Enterprise Application Integration (EAI) in connection with Business Intelligence solutions. However, as organizations move toward greater leverage of information assets by deploying Active Data Warehousing for tactical decision support, EAI becomes an important part of the architectural framework for a Business Intelligence solution. Data warehouse architects must work closely with corporate IT to ensure that a direction consistent with an organization's overall EAI framework is undertaken.

The Evolution of Decision Support Requirements

There was a time that data warehouse implementations were relatively isolated within the IT infrastructure of an organization. Data extract programs snuck out in the middle of the night or over a weekend to steal away with data from the operational systems and load it into the informational repository (data warehouse). A select group of knowledge workers (usually in marketing, finance, strategic planning, etc.) had access to this information for reporting and analysis, but the requirements for connectivity to large numbers of users outside the corporate ivory tower and interoperability with production systems were minimal.

With the expanded scope of successful data warehouses to encompass tactical decision making in addition to traditional decision support applications [1], the need for better integration into the mainstream of IT infrastructure has become essential. As a result, Enterprise Application Integration (EAI) has emerged as a critical component in the architecture of advanced data warehouse implementations. When properly deployed, EAI provides a vehicle for transitioning from back office decision making to tactical decision support on the front-lines to impact execution of the enterprise strategy. EAI plays a particularly important role when implementing the extreme data freshness service levels required for Active Data Warehousing.

What is Enterprise Application Integration?

From a business perspective, Enterprise Application Integration means providing unrestricted sharing of data and business processes among connected applications and data sources in the enterprise [2]. To realize this goal, architects must put into place a technical infrastructure capable of combining business processes, software and hardware platforms, and standards to allow seamless integration of two or more enterprise sys-

tems so that they operate as one (or at least provide the illusion of doing so to the business community). A variety of industry buzzwords all focus on this noble goal. Web services, message brokers, application servers, and middleware tools all provide aspects of the infrastructure necessary to realize the EAI vision. In this paper we will identify the technical components of an EAI deployment that are most relevant to a Business Intelligence solution.

The Role of EAI in Active Data Warehousing

An Active Data Warehouse deployment relies upon EAI infrastructure for both data acquisition and decision delivery. An Active Data Warehouse requires extremely up-to-date data from the transactional processing (bookkeeping) systems within an organization [3]. Advanced EAI infrastructure can facilitate (near) real-time data acquisition. The EAI infrastructure provides a bridge between the world of bookkeeping and decision making. When a business event is recorded in the bookkeeping systems the EAI infrastructure allows the decision-making environment (Active Data Warehouse) to become aware of the event on a real-time basis.

The bridge works the other way as well. When analytic applications in a tactical decision support implementation detect the need for an action, the EAI infrastructure is used to deliver decisions to the OLTP systems that will be responsible for the associated bookkeeping activities to make each proposed action a reality. EAI with process integration allows for "closed loop" decision making. Data fed from the bookkeeping environment into the Active Data Warehouse will (selectively) cause event-based triggers to fire (based on business rules) and initiate decisions that are fed back into the operational bookkeeping systems for execution.

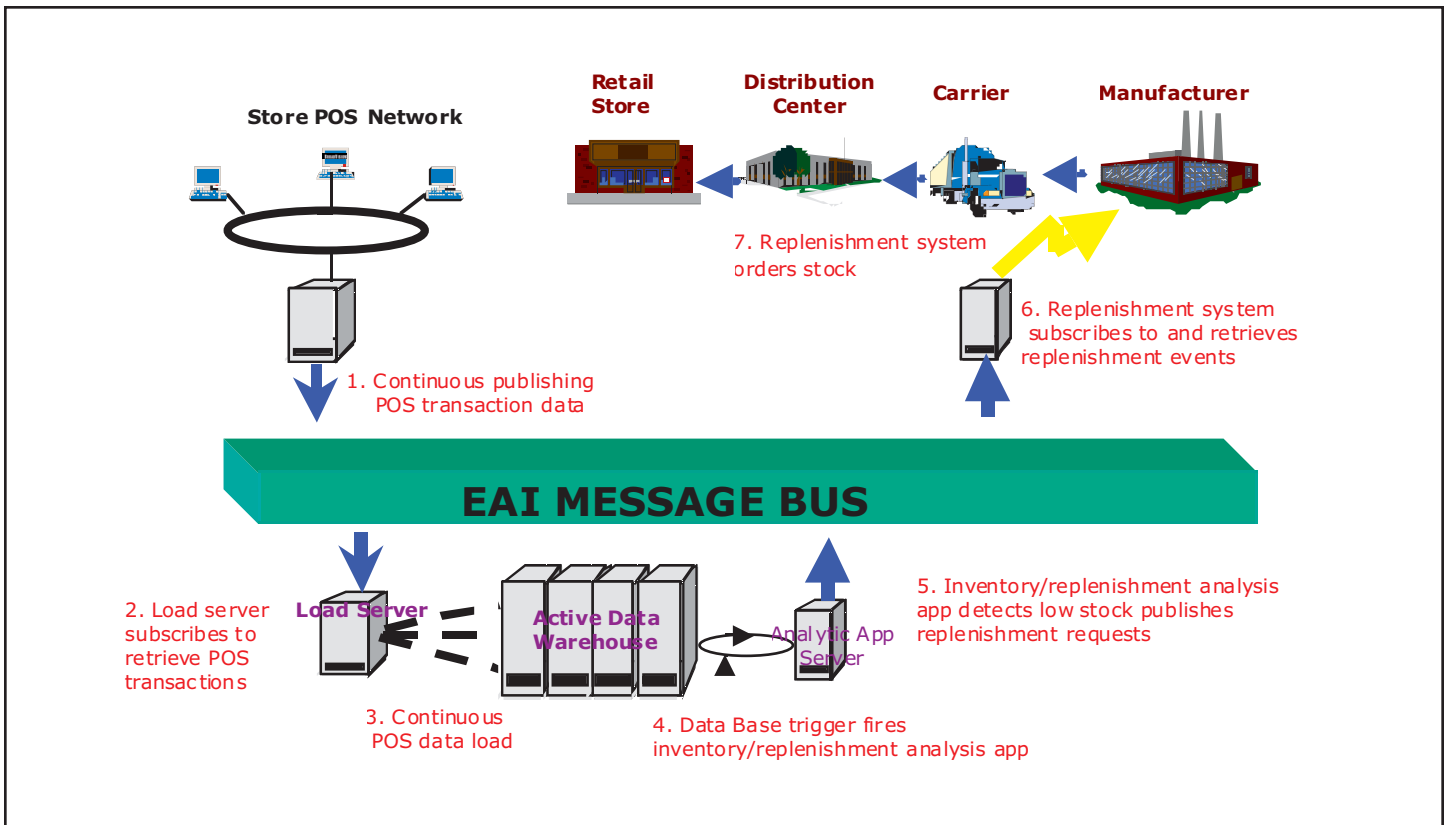


Figure 1: An Active Data Warehouse with Enterprise Application Integration

For example, consider the retail environment shown in Figure 1 where purchase transactions are captured using an electronic point-of-sales (ePOS) network in thousands of stores distributed across an extensive geography. Transactions from the ePOS systems are published to an EAI message bus as they occur in the stores. These business events are then delivered to all appropriate subscribers, including the Active Data Warehouse. Under certain conditions, such as when sales trends indicate a rapid depletion of inventory, the business rules embedded in the analytic capabilities of the Active Data Warehouse will arrive at a decision to order additional items for delivery to those stores that would otherwise end up with empty shelves. This inventory ordering decision is published using the EAI message bus and subscribers such as ERP (Enterprise Resource Planning) and General Ledger systems subscribe so as to be involved in the realization of the inventory re-order decision. EAI provides the “glue” to facilitate the closed loop cooperation between

the bookkeeping and decision-making systems.

An EAI Reference Architecture

Successful implementation of an EAI architecture involves much more than simply buying a packaged software solution. Realizing the full potential from an Active Data Warehouse deployment with EAI as a building block for data acquisition and delivery of decisions requires significant infrastructure. Some of the key components of the required reference architecture are as follows:

1. Secure, reliable, and scalable messaging: EAI implementation requires (near) real-time exchange of information among application components within the enterprise. The information exchanged is often business sensitive and must be protected from unauthorized acquisition or tampering. Moreover, reliable delivery of messages means that there is a guarantee that no message ever gets lost or delivered more than once. As more and more

of the enterprise becomes integrated into the EAI solution, the message delivery infrastructure must be able to incrementally grow bandwidth to handle increased messaging without degraded performance.

- 2. Intranet, Extranet, and/or Internet connectivity:** To exchange information requires enterprise connectivity. Intranet connectivity may be appropriate if sharing is strictly confined to the traditional enterprise. However, with virtual integration in the interest of greater cooperation between partnering organizations it is often desirable to re-think the traditional boundaries of the enterprise to allow inter-organizational connectivity or even customer interchange with the data warehouse using EAI technology. In such cases, Extranet and/or Internet connectivity may come into play.
- 3. Routing and transformation capability:** In an EAI implementation, information needs to get to

the correct recipients in a form appropriate to receiving applications. Routing capability in an EAI implementation ensures that messages are directed to all applications that need information about a business event. Transformation capability is required to convert message formats between applications so that seamless integration can take place without re-writing legacy application interfaces.

4. Publish/subscribe or point-to-point messaging model: Exchange of information between applications in an EAI implementation can be performed using either publish/subscribe or point-to-point messaging models. The publish/subscribe model is more general purpose in that it can easily support either direct application-to-application messaging or broadcast messaging to multiple subscribers. Point-to-point messaging can usually be implemented with lower overhead and greater simplicity than publish/subscribe messaging, but is best suited to direct application-to-application messaging. Theoretically, publish/subscribe can be simulated with point-to-point messaging, but practical considerations advise against doing so without a commercially packaged solution.

5. Non-intrusive application adapters: Enterprise application integration needs to be achieved with the least possible impact on legacy applications. Legacy applications are typically fragile and expensive to modify, so avoiding intrusion into applications being integrated is critical for overall success of the EAI endeavor. To facilitate this goal, a framework that allows applications to easily plug into the EAI infrastructure using interface adapters is required.

6. Business process management capability: There is much more to a successful EAI implementation than just pushing messages around between applications. A fully func-

tional EAI deployment supports workflow automation and directly contributes to business process management. Tools for defining workflow and automatically managing message routing and transformation using workflow meta data allows the EAI framework to be leveraged much more effectively.

The reference architecture described herein is not specific to the requirements of an Active Data Warehouse, but encompasses the necessary framework for EAI within an enterprise as a whole. The Active Data Warehouse is but one participant in the overall application environment within the enterprise. A data warehouse implementation will rarely dictate an organization's EAI requirements. Rather, when the data warehouse becomes "activated" for near real-time participation in organizational decision making, it will need to integrate with the existing (or emerging) EAI infrastructure of the corporation.

There are many different approaches and tools for EAI implementation. The Hurwitz taxonomy of the EAI implementations described in [4] provides a useful framework for understanding differences between the options available in the marketplace. From the perspective of Active Data Warehouse deployment, there are three categories of EAI that are most interesting:

1. Data-level EAI,
2. Message-level EAI, and
3. Process-level EAI.

The options listed here typically reflect an increasing level of maturity in the integration of the Business Intelligence environment with the overall enterprise application framework. In the paragraphs that follow, we describe each approach along with associated implementation tradeoffs.

Data-level EAI

The data-level EAI technique implements information exchange between multiple application data stores using traditional extract, transform, and load (ETL) techniques that are common-

place in data warehouse deployments. A variation on the ETL technique is ELT whereby data is extracted, loaded into staging tables, and then transformed inside the data warehouse platform. In either case, meta data is used to define transformations between the source system data content and the target data required by the data warehouse.

Data-level EAI is typically implemented using batch file processing techniques. A major advantage of data-level EAI is that it is extremely non-intrusive. As long as a window of time can be defined during which an extract from the operational source systems will not adversely impact production workload, the impact to legacy systems infrastructure should be minimal. No complex integration between systems is required; meta data specifications are used to translate between source data definitions and the target definitions for the data warehouse. The translation may be performed either within an engine-based tool or via code generated automatically from the meta data. In either case, all that the operational source systems need to provide are copy book definitions, data dictionaries, and other information used to understand source system content.

Tools that fall into the data-level EAI category include Ab Initio Co>Operating System, Ascential DataStage, ETT*Extract, Hummingbird Genio, Informatica PowerCenter, and many others.

Message-level EAI

The message-level EAI technique is much more amenable to (near) real-time sharing of information between source systems and an Active Data Warehouse. Message-level EAI manages the exchange of messages among multiple applications using reliable queuing systems. Business events can be published to one or more message queues as they occur in real-time rather than relying on a batch file processing model for information exchange.

The disadvantage of message-level EAI is that it requires more involvement from participating applications than data-level EAI. Participating applications in message-level EAI must create interfaces for sending and receiving messages. In some cases, there is a non-trivial amount of coding that must take place for a legacy application to implement message-level interfaces. This is specifically an issue with older, batch-oriented systems. It is difficult to leverage a real-time messaging capability if the source system is a batch billing system with no hooks for delivering information at a lower level than files.

Tools that fall into the message-level EAI category include IBM MQSeries, Microsoft MSMQ, and many others.

Process-level EAI

The process-level EAI technique goes beyond message-level EAI by overlaying a workflow management capability on top of message delivery capability. While message-level EAI is usually point-to-point, process-level EAI tools typically rely on a publish/subscribe messaging model. Process-level EAI can be thought of as an extension to message-level EAI with a middleware layer that performs the business process management using meta data derived from workflow automation tools.

Process-level EAI may be overkill for organizations just beginning with EAI implementation. However, full blown process-level EAI definitely emerges as a requirement as more sophistication is required for routing information between applications, orchestrating business processes for decision making and task execution, and interfacing to external business processes [5].

Tools that fall into the process-level EAI category include SeeBeyond Business Integration Suite, TIBCO ActiveEnterprise, Vitria BusinessWare, webMethods Integration Platform, and many others.

Marketplace Trends in EAI

More advanced EAI deployments are clearly emerging as a requirement in the Business Intelligence domain as data warehouses evolve from passive to active implementations. The ability to acquire data in near real-time and deliver decisions back to bookkeeping systems for operational execution requires much tighter integration of the data warehouse into the enterprise computing environment to enable and leverage tactical decision support capabilities.

Interestingly, however, is that message-level and process-level EAI tools typically fall short of data-level EAI tools in handling complex data transformations. For example, most process-level EAI tools can handle “scalar” transformations between data types or domain definitions, but not transformations requiring many-to-one record mappings or access to external data sources as part of the transformation rule execution. Traditional ETL/ELT tools are much more sophisticated in their meta data driven transformation capabilities than the message-level and process-level EAI tools.

As a result, message-level or process-level EAI tools are often used to feed into data-level EAI tools for transformation of data before delivery into an Active Data Warehouse. The message-level or process-level EAI tools provide reliable, near real-time delivery of information with final data transformation and loading implemented with traditional ETL tools. However, this approach only works if the ETL tool can handle continuous data acquisition, transformation, and loading (as opposed to file-based processing for these tasks). Leading innovators in the ETL marketplace have begun implementing stream-based interfaces for handling near real-time data delivery to augment their traditional file-based interfaces used for batch processing. Moreover, EAI adapters that allow data to be acquired into ETL tools for transformation processing from message-level and process-level EAI tools have begun to emerge into the marketplace. In the future, we expect a conver-

gence between data-level, message-level, and process-level technologies in the EAI marketplace as vendors consolidate capabilities.

Conclusions

Organizations are finding that it is high time that they remove their data warehouse assets from “solitary confinement” in the corporate ivory tower. Leveraging information assets in an Active Data Warehouse in support of tactical decision making requires that the Business Intelligence solution within an enterprise cooperate more closely with OLTP applications to facilitate near real-time data acquisition as well as delivery of decisions. Enterprise Application Integration (EAI) plays a critical role in providing seamless integration of decision-making capability with traditional bookkeeping systems within an enterprise. The specific tools used for EAI deployment will depend on maturity of the technical infrastructure and sophistication of requirements within the enterprise.

- [1] Brobst, S. and J. Rarey. Leveraging an Enterprise Data Warehouse for Tactical Decision Support. *Data Warehousing Institute Flashpoint*. May 25, 2001.
- [2] Linthicum, D. *Enterprise Application Integration*. Addison-Wesley. 1999.
- [3] Brobst, S. Delivering Extreme Data Freshness with Active Data Warehousing. *Journal of Data Warehousing*. The Data Warehousing Institute. Spring, 2002.
- [4] Gold-Bernstein, B. EAI Market Segmentation. *EAI Journal*. July/August, 1999.
- [5] Schmidt, J. Enabling Next Generation Enterprises. *EAI Journal*. July 1, 2000.

Brobst, S. “Active Data Warehousing and Enterprise Application Integration,” *Proceedings of Data Warehousing 2002: From Data Warehousing to the Corporate Knowledge Center*, Physica-Verlag Heidelberg, November 12-13, 2002. pp. 15-23.