



claudia
imhoff

Reprinted from
DMReview

Understanding the Three E's of Integration EAI, EII and ETL

Remember the story of Humpty Dumpty? All the King's horses and all the King's men couldn't put Humpty together again. Many IT professionals tell me that they often feel like they live in Humpty Dumpty land. Ever since the first two computer programs were written, IT has struggled with the resulting disintegration - putting the data and applications together again. Integration of data and applications across the enterprise has been the long-standing goal of many organizations; however, until recently, we have been limited in the technological help to achieve this goal.

Fortunately, we have three technologies to help with this. I call them the three E's - enterprise application integration (EAI), enterprise information integration (EII) and extract, transform and load (ETL) software.

These technologies range from the need for real time versus batch integration and from the need for the integration of data versus the integration of applications. Figure 1 demonstrates where the three technologies reside in terms of these two spectra. If your needs are for real time data integration, then EII is the best bet. If you need batch data integration, then ETL is a better fit. And, if your need is for batch or real time application integration, EAI is the most appropriate tool.

With any new technologies such as these, there are always substantial amounts of confusion regarding what each technology really does and when it should be used. To avoid this, you need to develop a clear definition of these technologies and then determine where and when you would use them. Let's start with definitions. After

talking to a number of vendors such as Composite Software and Celequest, and to several clients, I came up with the following:

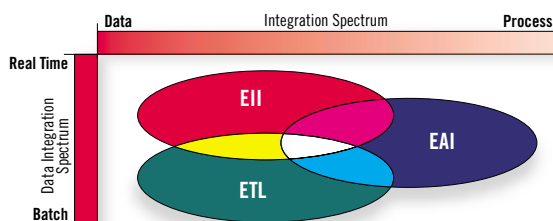
- EAI: a framework by which an organization centralizes and optimizes application integration, usually through some form of push technology that is event-driven. The target for this technology is usually an application.

With any new technologies, there are always substantial amounts of confusion regarding what each technology really does and when it should be used.

- ETL: a framework that assimilates data, usually through batch processing, from the operational environment of heterogeneous technologies into integrated, consistent data suitable for consumption by the decision support processes. The target for ETL technology is a database such as a data warehouse, data mart or operational data store.

- EII: a framework for real-time integration of disparate data types from multiple sources inside and outside an enterprise, providing a universal data access layer, using pull technology or on-demand capabilities. The target for EII is a person, via a dashboard or a report.

Let's focus on where each of these technologies fits into your architecture. Figure 2 shows the best place to use each of the three E's. EAI integrates transactions between two or more applications; ETL integrates data between your operational systems and your decision



Copyright © 2005, Intelligent Solutions, Inc., All Rights Reserved

Figure 1: The Integration Landscape Today

support components; EII creates virtual data integration between various sources of data.

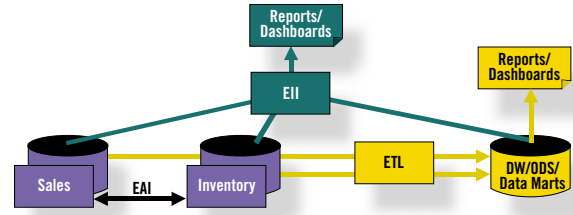
For you to use these technologies optimally, you must understand where each one is best positioned to be most practical.

EAI is most useful when you need to connect applications in real time for business process automation. Another practical use for EAI is in making a change (typically to a small set of records) in one application and reflecting it elsewhere in other applications. This technology is very good at ensuring that the change is captured and delivered reliably to the appropriate application or system.

You will find ETL to be most useful when you need to produce a data warehouse of well-documented and reliable data for historical analyses such as time series analysis or multidimensional queries. The tool is also used to integrate key master data. ETL shines for activities such as removing duplicate data, invoking data quality processes and so on. These tools are also used to build discrete data marts to serve a functional or departmental area and to serve a unique long-term purpose. ETL tools allow the implementer to put a repeatable process in place for consistency and reusability, which includes the creation of accurate technical meta data, supporting the overall integrity of the business intelligence (BI) environment.

EII is most useful when you need to create a common gateway with one access point and one access language to disparate data sources. These tools provide more flexible and ad hoc access to data by end users or applications without requiring permanence or a long-term purpose. They are able to access XML, LDAP, flat files and other non-relational data in addition to traditional relational databases, and they can publish relational data as XML/Web services data. EII is particularly useful in supplementing master data warehouse (DW) data with additional or real-time detail (e.g., combining historical data with the current situation).

In addition to understanding these cases of when to use these technologies, you should also understand some challenges that go along with all of them. First, they require that your implementers have a thorough understanding of the data requirements for both strategic and tactical decision making. With ETL, this ensures that the appropriate data is extracted, transformed and loaded, ready for



Copyright © 2005, Intelligent Solutions, Inc., All Rights Reserved

Figure 2: Where EAI, EII and ETL Fit Into Your Architecture

use by the analysts directly or for consumption by an EII server. With EII, it ensures that the views you design and build meet the analysts' reporting requirements. In all cases, understanding your data sources and requirements is a necessary step and is worth the significant time it can take.

It also must be recognized that bringing these tools into your overall architecture requires a commitment from both the business and IT to develop a data and application management strategy that creates an ongoing process. Part of this strategy must be the recognition that your archiving mechanisms become quite important and that audit trails must be established from the start. These are needed to ensure consistency and reliability of the integrated data or applications.

Finally, it is important to constantly monitor the performance and efficiency of these technologies in your particular infrastructure. Their performance will be greatly influenced by the archive duration, data size and granularity, and overall load performances. Performance also includes the impact these tools may have on your operational applications and systems. Be sure you constantly monitor what, if any, impact they have on these systems. ^{DMR}

Claudia Imhoff, Ph.D., is the president and founder of Intelligent Solutions (www.intelsols.com), a leading consultancy on CRM and business intelligence technologies and strategies. She is a popular speaker and internationally recognized expert and serves as an advisor to many corporations, universities and leading technology companies. She has coauthored five books and more than 50 articles on these topics. She may be reached at cimhoff@intelsols.com.