

Controlling Your Data 1—Strategies

by Kevin Pledge

This is the first of two articles discussing technologies used to control data. This article covers data management strategies, while the second article will cover extraction, transformation and loading tools, and the technology most likely to be used for analytical applications in insurance businesses.

What Do We Mean by “Controlling Data?”

The primary purpose of data management technology is to deliver high-quality data. By this we generally mean:

- data that can be verified against its source,
- a single agreed-upon definition for the data, and
- being in compliance with data standards

In addition, there are a number of secondary goals that must also be met by any tool used to control data. It must be able to handle large volumes of data and move it efficiently. The system must maintain the security of the data. Processes must be verifiable and support audit and testing. Finally, depending on the environment and application, there may be some additional requirements, such as the ability to handle data from multiple sources, multiple platforms or in real time.

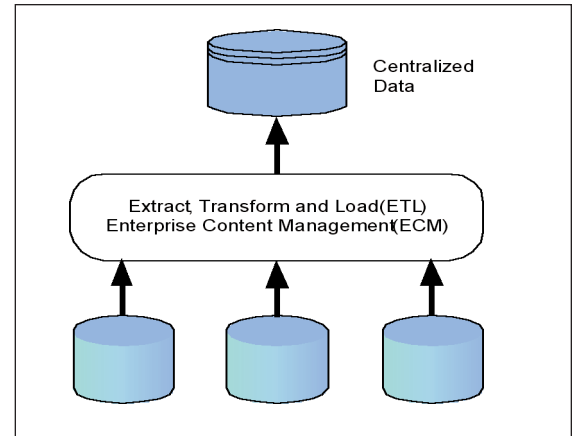
There are three data management approaches:

- consolidation
- federation
- propagation

Consolidation

Consolidation involves moving data from multiple sources to a single, consolidated source. The two types of tools designed for this are enterprise content management (ECM) tools and extraction, transformation and loading (ETL) tools. ECM is typically used for knowledge management, such as organizing and accessing information from documents, while ETL is used for data.

Consolidation is typically very scalable and can support complex data transformation. Its main weakness is that the data is not available in real time.



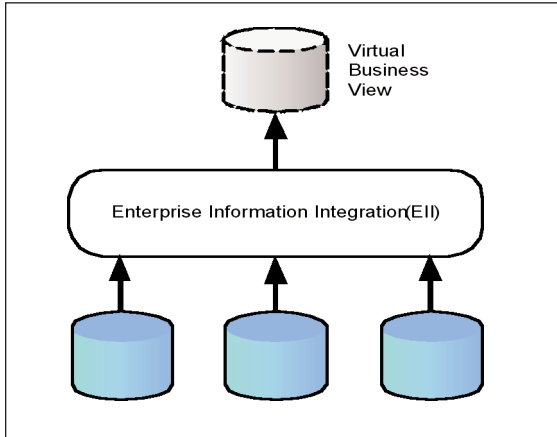
Federation

Federation involves creating a virtual business view of the primary data sources. These tools are typically referred to as enterprise information integration (EII).



Kevin Pledge, FIA, FSA, is president and CEO of Insight Decision Solutions in Markham, Ontario. He can be contacted at kpledge@insightdecision.com.

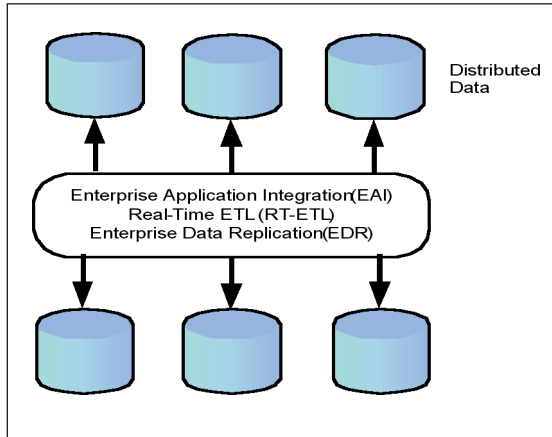
As a virtual view of the business, any changes in data are reflected in the virtual-business view in real time. The downside of this is that data cleansing and transformation is limited.



Propagation

Propagation tools distribute and replicate or copy data between multiple systems, often in real time. These tools may be referred to as enterprise application integration (EAI), real-time ETL (RT-ETL) or enterprise data replication (EDR).

Once again, the main advantage is that data can be available in real time, however performance can vary and the tools may be difficult to set up.



In Summary

There are three fundamental approaches for data control—consolidation, federation and propagation. Federation and propagation can share data in real time, but are limited in their ability to apply complex transformations and manage large volumes of data. For this reason a consolidation strategy is of most relevance for data management for actuarial applications. ETL tools are the topic of the next article in this series. 🖥️

