

Whitepaper

Solving Complex Hierarchical Data Integration Issues

What is Complex Data?

Historically, data integration and warehousing has consisted of flat or structured data that typically comes from structured operational applications and databases.

Data integration today is becoming a much more complex undertaking that involves a variety of platforms, formats, and data sources. Incoming data often has hierarchical structures that must be parsed into a common format, validated, and loaded into internal databases. And for outbound data the challenges are even greater because data from a number of disparate sources such as databases, files, and services must be combined in a structure that complies with established standards and partners' requirements.

Data integration solutions are being developed to acquire, merge, and transport data in a myriad of formats. Integrated data is loaded into target databases and applications inside the organization or packaged into files and documents for exchange with other organizations. Growth has expanded both analytic data integration in the form of extract, transform, and load [ETL] for data warehousing and operational data integration for consolidation, migration, or synchronization of operational databases. Despite the development of sophisticated data integration solutions in recent years, few adequately address the integration and management of complex hierarchical data.

Types of Data

- Structured Data - Simple tabular data structures from relational database management systems, tables, and flat files in record format.
- Unstructured Data - Documents of mostly natural language text, like word-processing files, email, and text fields from databases and applications. Some documents may have light metadata, such as spreadsheets and RSS feeds.
- Semi-structured Data - Data documents exchanged between organizations often combine unstructured and structured data, or, such as in XML, text that has been structured with metadata tags. Semi-structured data documents often comply with open data exchange standards such as SWIFT, NACHA, HIPAA, HL7, RosettaNet, and EDI.

The data processing capabilities of most data integration solutions addresses structured data sources like tabular data from relational database management systems (RDBMSs) and flat files in record format. Unfortunately, much of an organization's data these days is not structured in a tabular way, limiting the benefits that are derived from its data gathering efforts to tabular data from structured sources, even though similar value and benefits could also be achieved by embracing the complex data found in unstructured, semistructured, and hierarchical data sources. As data integration practices and technologies expand to embrace complex data, data integration solutions must grapple with two tasks that are new to most data integration specialists: integrating data from complex and nontraditional sources and assuring the quality of data drawn from those sources.



Complex Data Integration for Business Users

Building hierarchical documents from relational or flat data is a complex undertaking. It is usually accomplished using hand-coded programs developed by IT professionals that are time consuming to write, expensive to maintain, and don't always meet the needs of end users. Businesses are beginning to recognize that business analysts and those who make business decisions should be developing data integration processes to meet their guidelines rather than handing off to developers whose expertise is in perfecting systems but not in how the data from the systems is used.

Approachable Data Integration — The good news is that there is a powerful yet affordable complex data integration software solution on the market that doesn't require difficult and expensive IT coding processes. Astera's Centerprise Data Integrator has been developed from the ground up specifically so business users can solve their complex data integration issues with XML, electronic data interchange (EDI), web services, etc. first hand. The intuitive, drag-and-drop environment enables non-developers to quickly and easily parse and construct hierarchical structures and manage their complex integration jobs faster and more affordably. Centerprise approachable data integration significantly lowers the need for IT resources and empowers those who use the data to have a say in how it is processed.

Working with Hierarchical Data

The key to successful management of complex, hierarchical data is superior data mapping functionality. Data mapping is very difficult and must be done correctly in order to deliver the desired results. New technologies have been created in Centerprise that simplify and automate the complexities of data mapping for business users. Disparate data sources can be joined via the drag-and-drop environment to build complex trees, automatically build hierarchies from relational databases, and apply transformations to specific tree nodes. With these unique capabilities, users can rapidly create powerful transformations for hierarchical-to-flat, flat-to-hierarchical, and hierarchical-to-hierarchical situations.

Single, integrated platform — Centerprise provides an integrated and consistent approach for working with both flat and hierarchical files. Because everything is done in a single environment, users working with hierarchical data use the same familiar toolset they use when working with flat structures. All data parsing and structure building is done in the same map,

with no need for coding or scripting.

Special components and modes — Technologies have been created specifically to provide capabilities for joining disparate data sources to build complex trees, automatically building hierarchies from relational databases, and applying transformations to specific tree nodes. This last item—applying transformations to tree nodes—enables users to apply common Centerprise transformations such as Aggregate, Filter, Join, Distinct, Sort, Normalize, Denormalize, and others to specific nodes in a hierarchical structure. With this unique technology, users can rapidly create powerful transformations for hierarchical-to-flat, flat-to-hierarchical, and hierarchical-to-hierarchical situations.

Hierarchical formats and databases

Centerprise supports most hierarchical source formats from databases, files, and web services. Database hierarchies include data models and visual joins. Data models are easy to create in Centerprise for joins that happen in the database. A new and more flexible way to create a data model is through a “visual” join. Typical hierarchical file formats include COBOL, EDI, XML, reports/PDF files, delimited, and fixed length.

EDI - Electronic Data Interchange

Today’s agile business relies on EDI technology to automatically exchange business documents from different sources in a standard electronic format, thus streamlining operational costs, increasing processing speed, reducing errors, and improving relationships with business partners. Centerprise provides complete bidirectional EDI mapping functionality for creating customized transaction structures, parsing and validating incoming documents, building outgoing documents, and creating outgoing EDI files, including HIPAA-compliant documents.

With Centerprise EDI you can:

- Create custom EDI formats using the Centerprise EDI Message Designer
- Parse and validate incoming EDI files. Validations include data type, lookup codes, message, functional group and interchange integrity, and conditionals. Error information can also be mapped to create customized outputs.
- Send and receive messages in all current and past versions of the X12 standard. Additional dialects will be available in the third quarter of 2014.
- Build standards and partners’ compliant outbound EDI files

Web Services

Input and output payloads are almost always hierarchical. Centerprise supports web services such as SOAP and REST, as well as XML and JSON. The software also provides a gateway for countless consumer and enterprise APIs.

- Centerprise XML functionality can be used to parse incoming documents as well as build outgoing documents. Key functionalities include:
 - Use an existing schema or build one by providing a sample XML or JSON file
 - High-performance reading and writing capabilities able to work efficiently with very large data files
 - Quickly build complex outgoing XML documents using multi-table queries, tree joins, scoped transformations, and other features

Custom Hierarchical Files

Centerprise also enables parsing and creation of custom hierarchical files. This feature can be used to process hierarchical

delimited files produced by many systems.

Centerprise Object Model

The Centerprise object model now includes a layout tree to organize referenced objects versus collections, as well as elements. Data preview is different for hierarchical data than for flat files and introduces the concept of anchors and cardinality rules.

Hierarchical Transformations

Centerprise brings an innovative approach to building complex hierarchical documents that uses a combination of multi-table queries, tree-joins, lookups, and sort, aggregate, and other transformations to progressively build output using a visual, stepwise method. At each step, you can use Centerprise's single-click preview to see data in at that step, making it easy to identify and correct any mapping errors.

Centerprise not only provides the functionality to parse incoming EDI, XML, and custom hierarchical files, it also offers an extensive array of unique transformations that can be used to transform files into hierarchical or flat structures, including Filter, Sort, Route, Union, Tree Join, and Pass-through. Loading hierarchical data into a database requires sequencing of inserts or updates to ensure that any referential integrity constraints are not violated. For instance, a purchase order must be inserted before line items can be inserted. Centerprise features the ability to automatically sequence database writes based on referential constraints.

Supplemental Transformations

Centerprise provides some unique transformations that are very useful when dealing with trees, including FLWOR, Scoped Transformations, Aggregate Transformations, and text processors.

- FLWOR (FOR, LET, WHERE, ORDER BY, RETURN) must be scoped and allows for querying the tree for a specific node.
- Scoped Transformations are created with a "Scope Map" and are a must when dealing with sub-trees. You virtually "attach" a new sibling node to the existing tree.
- Aggregate Transformations pull aggregate data from sub-collections and must be used with a scope
- Text processors deal with two formats at once and are ideal for scenarios where database columns contain structured text (i.e., xml)

Common Scenarios

Want to learn more? Our one-hour webinar, Complex Data Integration in Centerprise, is available for viewing on our Astera.TV video library portal. Along with demos of all the technologies discussed in this blog, the video contains several informative common scenario demos:

- Creating hierarchies from multiple flat sources
- Flattening hierarchies
- Writing to relational tables
- One hierarchy to another

Centerprise Solutions



Data Warehousing

High-performance data warehousing ETL features in a unified, intuitive environment.



CDC

Choose either a batch or real-time change data capture strategy for your particular requirements.



Data Mapping

Transforms advanced mapping, validating, and cleansing tasks into basic drag-and-drop or single-click commands.



ETL

Extract data from any source, transform it to suit your needs, and load it into your database or warehouse.



Data Conversion

Complex doesn't need to be complicated. Visual, code-free parsing, transforming, and loading of data from any source.



Data Integration

A single platform for complex, hierarchical integration the requires no coding.



Data Migration

Unique hierarchical data processing technologies automate and streamline data migration projects.



EDI

Full electronic data interchange functionality combined with Centerprise complex data mapping capabilities.



www.astera.com

Contact us for more information or to request a free trial
sales@astera.com 888-77-ASTERA

Copyright © 2014 Astera Software Incorporated. All rights reserved. Astera and Centerprise are registered trademarks of Astera Software Incorporated in the United States and / or other countries.
Other marks are the property of their respective owners.