

Innovations in Workers Compensation: XML Shredding for External Data Integration

Devidas Kanchetti¹ Rajesh Munirathnam² and Darshit Thakkar³

¹Corresponding Author

¹Data Scientist, 3808 Fox Run Rd, Matthews, North Carolina 28104 USA

Email: kanchetty@gmail.com

²Data Scientist, 872 Robin Road, Hillsborough, New Jersey 08844 USA

Email: mrajesh.qs@gmail.com

³Product Manager - Technical, 3920 Mystic Valley Pkwy, 818, Medford, Massachusetts 02155 USA

Email: darshit.thakkar07@gmail.com

Abstract

In the rapidly evolving landscape of data management, the integration of external vendor data has become a pivotal challenge for organizations aiming to streamline operations and enhance decision-making processes. This journal, "The Transformative Power of XML Shredding - External Data Integration," explores the intricate methodologies, frameworks, and innovations in XML shredding, emphasizing its critical role in integrating diverse data sources. XML shredding, the process of converting XML documents into relational or other structured formats, offers a robust solution for harmonizing disparate data sets, ensuring compatibility and coherence across various platforms and systems.

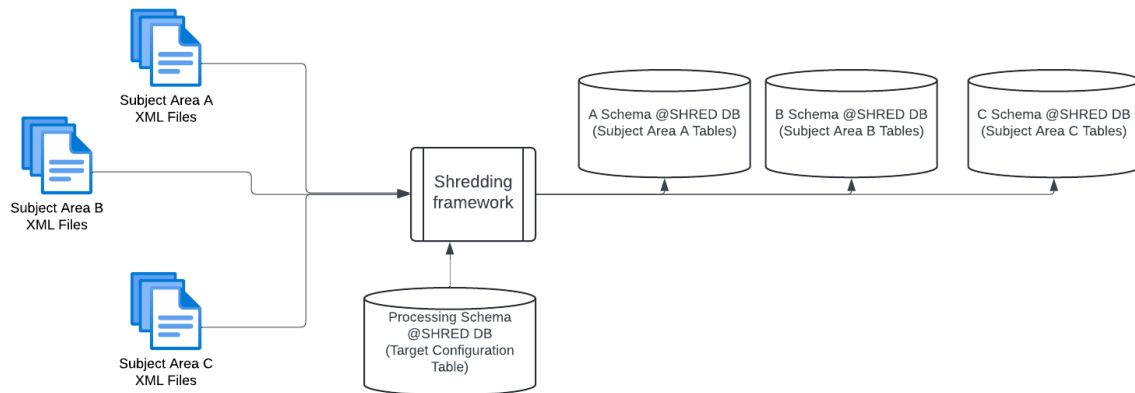
Keywords: Innovations in Workers Compensation, XML Shredding, External Data Integration

1. Introduction

In today's data-driven world, the effective management and integration of diverse data sources pose significant challenges to organizations across industries. The advent of XML shredding frameworks has emerged as a pivotal solution in addressing these challenges, offering structured methodologies and tools to streamline data transformation and integration processes.

This journal, dedicated to exploring XML shredding frameworks, aims to delve into the intricate details and advancements shaping the landscape of data integration. XML shredding, a technique that converts XML documents into relational or other structured formats, plays a crucial role in harmonizing disparate data sources. By breaking down complex XML structures into manageable components, these frameworks enable seamless integration with existing systems, enhancing data interoperability and accessibility.

2. Process Diagram



3. Use case

Insurance firms sourcing data from a vendor product find that vendors are unwilling to grant database access, opting instead to provide data in XML documents. These insurance firms require the XML data for their data analytics solutions, necessitating extraction and transformation into relational tables.

4. Advantages

XML shredding offers businesses a powerful and efficient method for managing data in third-party application integrations, improving interoperability, data consistency, performance, scalability, security, and compliance.

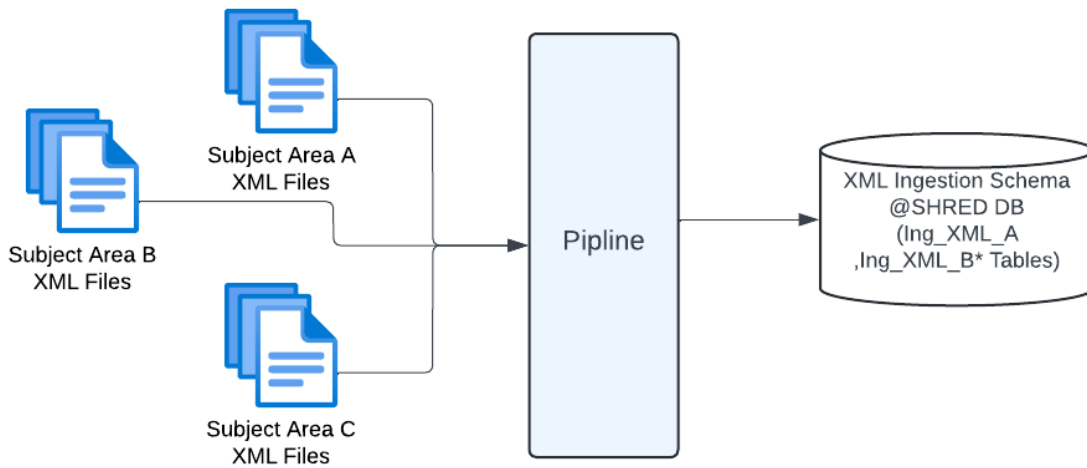
5. Risks and Impact on Application

In case the vendor adds additional required data fields, the extraction and processing process will need to be modified accordingly to accommodate these changes.

6. Tiers

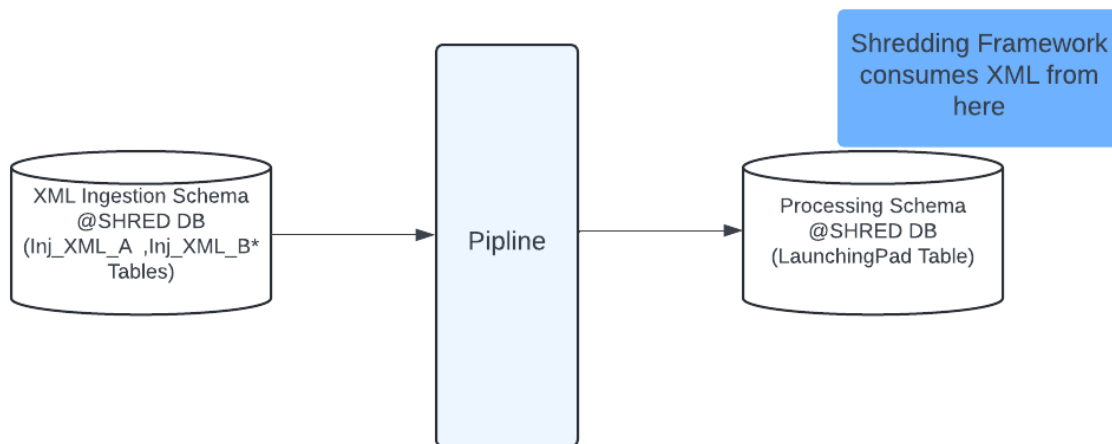
Ingestion Layer/ Tier

This batch process retrieves files from the vendor site to our location and attaches the XML files into relational table columns. This ingestion of XML files makes the files available in the relational table for further processing.

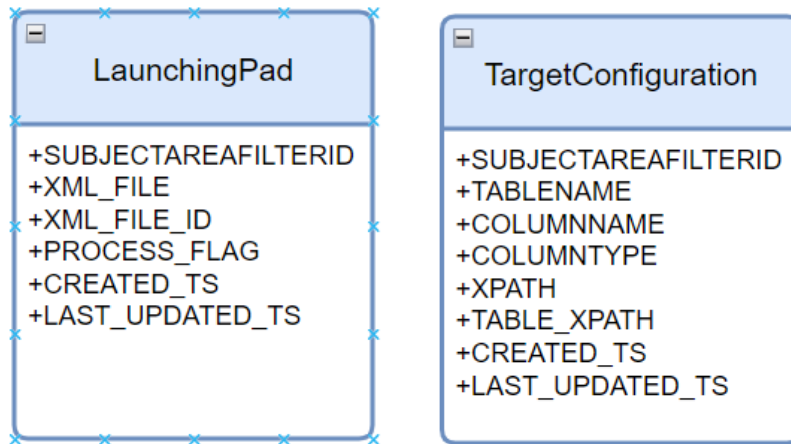


Stage Layer/ Tier

In this layer, the pipeline loads data from source tables into the launching pad table. The launching pad table is crucial as it serves as the main repository for all attached XML files. Each entry in the launching pad table includes a sequence number, file ID, and a process flag. As the control program processes each file (record) in the launching pad table, the process flag is updated. For each file, the root tag is red and loaded into a column in LaunchingPad table column File_Id.



Launching Pad, Target Configuration Tables Structures and Sample Data

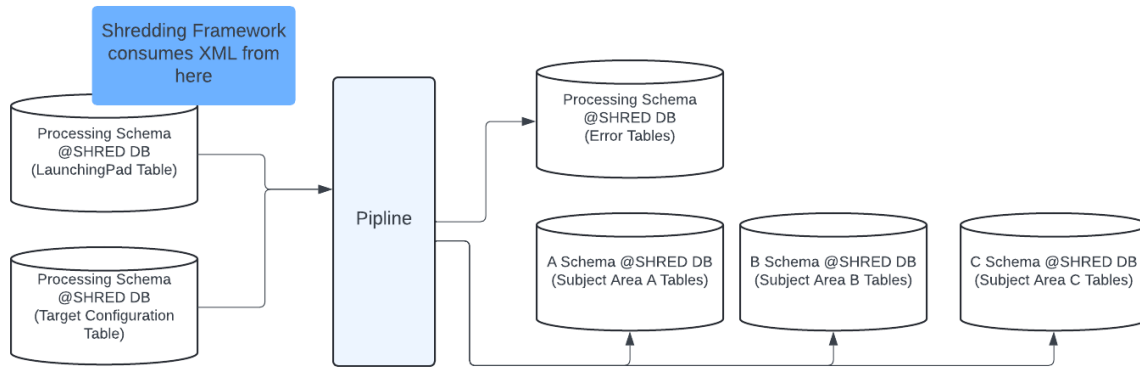


SUBJECT AREA FILTERID	XML_FILE	XML_FILE_ID	PROCESS_FLAG	CREATED_TS	LAST_UPDATED_TS
110	<<XML>>	12345	1	07/17/2001 12:22:10	07/17/2001 12:22:10
120	<<XML>>	2345	0	07/16/2001 12:22:10	07/16/2001 12:22:10

SUBJECT AREA FILTERID	TABLENAME	COLUMNNAME	COLUMNTYPE	XPATH	TABLE_XPATH
110	PRODUCT	TYPE	VARCHAR2(50)	/info/kali[@index="4"]	/info
110	PRODUCT	NAME	VARCHAR2(100)	/info/kali[@index="5"]	/info
120	PROJECT	NAME	VARCHAR2(100)	/osa2/van[@index="1"]	/osa2
120	PROJECT	LOCATION	VARCHAR2(100)	/osa2/van[@index="2"]	/osa2

Process Layer/ Tier

In this layer, the XML shredding control process or pipeline reads XML files from the Launching pad table and processes them according to configuration information in the Target Configuration table. It loads data into relevant target tables across different subject areas. As each XML file is processed, the process flag in the launching pad table is updated.



7. Conclusion

In conclusion, the adoption of XML shredding represents a significant advancement in data management practices, particularly in the context of integrating external vendor data. By enabling the transformation of XML documents into structured formats, organizations can effectively overcome the challenge of disparate data sources. This process not only enhances operational efficiency but also strengthens decision-making capabilities by ensuring data compatibility and coherence across different platforms and systems. Moving forward, further research and implementation of XML shredding methodologies promise to continue driving innovation and efficiency in data integration practices, supporting organizations in achieving their strategic objectives in an increasingly complex data landscape.

7. References

- Beeri, C., & Tzaban, Y. (1999). SAL: An algebra for semi-structured data and XML. In Proc. of the 2nd Workshop on the Web and Databases (pp. 37–42). Philadelphia.
- Bohannon, P., Freire, J., Roy, P., & Simeon, J. (2002). From XML schema to relations: A cost-based approach to XML storage. In Proc. of the 19th Int. Conf. on Data Engineering (pp. 64–75). San Jose.
- Carey, M., Florescu, D., Ives, Z., Lu, Y., Shanmungasundaram, J., Shekita, D., & Subramanian, S. (2000). XPERANTO: Publishing object-relational data as XML. In Proc. of the 3rd Workshop on the Web and Databases. Dallas, Texas.
- Christophides, V., Cluet, S., & Simeon, J. (2000). On wrapping query languages and efficient XML integration. In Proc. of the ACM SIGMOD Int. Conf. on Management of Data (pp. 141–152). Dallas, Texas.
- Dayal, U. (1987). Of nests and trees: A unified approach to processing queries that contain nested subqueries, aggregates, and quantifiers. In Proc. of the 13th Int. Conf. on Very Large Data Bases (pp. 197–208). Brighton, England.
- DeHaan, D., Toman, D., Consens, M. P., & Özsu, M. T. (2003). A comprehensive XQuery to SQL translation using dynamic interval encoding. In Proc. of the ACM SIGMOD Int. Conf. on Management of Data (pp. 587–598). San Diego, CA.

- EI-Sayed, M., Dimitrova, K., & Rundensteiner, E. A. (2003). Efficiently supporting order in XML query processing. In Proc. of the fifth ACM Workshop on Web Information and Data Management (WIDM) (pp. 147–154). New Orleans, Louisiana.
- Fernandez, M., Tan, W. C., & Suciu, D. (2000). SilkRoute: Trading between relational and XML. In Proc. of the 9th Int. World Wide Web Conf. (pp. 16–25). Amsterdam, Netherlands.
- Fernandez, M., Simeon, J., & Wadler, P. (2000). An algebra for XML query. In Proc. of the 12th Conf. on the Foundations of Software Technology and Theoretical Computer Science (pp. 116–128). Delhi.
- Fiebig, T., & Moerkotte, G. (2001). Algebraic XML construction and its optimization in Natix. *World Wide Web Journal*, 4(3), 167–187.
- Florescu, D., & Kossmann, D. (1999). Storing and querying XML data using an RDBMS. *IEEE Data Engineering Bulletin*, 22(3), 27–34.
- Gonnet, G. H., & Tompa, F. W. (1987). Mind your grammar: A new approach to modeling text. In Proc. of the 13th Int. Conf. on Very Large Data Bases (pp. 339–346). Brighton, England.
- Kadiyska, Y., & Suciu, D. (2004). Mixed XML/relational data processing. In Informal Proc. of the workshop on programming language technologies for XML (PLAN-X 2004) (pp. 73–82). Venice, Italy.
- Kanne, C. C., & Moerkotte, G. (2000). Efficient storage of XML data. In *Proc. of the 17th Int. Conf. on Data Engineering* (p. 198). San Diego, CA.
- Manolescu, I., Florescu, D., & Kossmann, D. (2001). Answering XML queries on heterogeneous data sources. In Proc. of the 27th Int. Conf. on Very Large Data Bases (pp. 241–250). Rome.
- Paulley, G. N. (2000). Exploiting functional dependence in query optimization. PhD thesis, Dept. of Computer Science, University of Waterloo, Ontario, Canada.
- Shanmugasundaram, J., Tufte, K., He, G., Zhang, C., DeWitt, D., & Naughton, J. (1999). Relational databases for querying XML documents: Limitations and opportunities. In *Proc. of the 25th Int. Conf. on Very Large Data Bases* (pp. 302–314). Edinburgh, Scotland.
- Shanmugasundaram, J., Kiernan, J., Shekita, E., Fan, C., & Funderburk, J. (2001). Querying XML views of relational data. In Proc. of the 27th Int. Conf. on Very Large Data Bases (pp. 261–270). Rome.
- Wang, X., Luan, J., & Dong, Y. (2003). An adaptive and adjustable mapping from XML data to tables in RDB. In Efficiency and Effectiveness of XML Tools and Techniques and Data Integration over the Web VLDB 2002 Workshop EEXTT and CAiSE 2002 Workshop DIWeb (pp. 117–130). Springer-Verlag Berlin Heidelberg.
- Zhang, H. (2003). XML query processing and optimization. PhD thesis, School of Computer Science, University of Waterloo, Ontario, Canada.

Zhang, H., & Tompa, F. W. (2003). XQuery rewriting at the relational algebra level. In Trends in XML technology for the global information infrastructure (special issue of Int. Journal of Computer Systems Science and Engineering, 18(5), 241–262). September.

Author Biography

Devidas Kanchetti

A seasoned Senior Data Engineer adept in data modeling, advanced SQL, and ETL development.

Rajesh Munirathnam

Experienced Senior Information Technology Leader with a proven track record of driving innovative and results-driven strategies in the realm of data and analytics implementations for high-availability applications. Possesses a strong acumen for managing fast-paced environments and navigating complex technological landscapes. Known for delivering impactful solutions that enhance operational efficiency, drive business growth, and elevate organizational performance.

Darshit Thakkar

Distinguished product management professional with over 12 years of extensive experience in software industry targeting the business and consumer markets; consistently demonstrated the ability to manage the product life-cycle from concept to launch in a fast-paced environment with focus on machine learning and artificial intelligence products