Significance of Metadata and Data Modeling of Metadata by using Mark Logic

K. Lakshmi Prasanna, Jangala. Sasi Kiran, K Sreerama Murthy

Abstract: Metadata means data about data which illustrates, traces, and is simple to locate a resource. Metadata has been impacting almost every firm. It has become mandatory for organizations to know the data flow across business processes to take strategic decisions. But, collecting metadata across departments/business processes and putting into a commonality is very difficult by using conventional databases. We need to concentrate on the metadata managing technologies. There are data models that are designed which work on NoSQL database. Envelope pattern in Marklogic provide commonality for the metadata across processes. The data that is gathered across different processes need to be managed in a consistent way. We want to verify metadata management in banking domain. In this paper, we have ingested metadata across multiple departments in banking domain and verified the performance of search results.

Index Terms: Metadata; Data modeling; Mark logic; envelope pattern.

I. INTRODUCTION

A key development is bringing a data strategy that precedes metadata management. The available resources and business data can be reused at optimum extent and it can generate safety reports at regular intervals, risk compliance reports etc. Metadata generates profiles of key assets and can reuse real world data base based on its impact. Content retrieval for strategic decision making is simple with metadata by classifying key elements like document type, author, document created time stamp etc. With metadata one can improve different content types alerting with a key element such as creation or submission time stamp, in this case, one need not require actual content; the metadata will suffice the need. Firms even need to submit risk approvals and need to generate multi functional reports. Technically metadata can be defined as running complicated commands by considering the strong search indexes without the need for the original data. The metadata indexes will make the data framework as a search engine where one can search over data and metadata. As per Data versity report on Metadata, around one third of firms have begun dealing with metadata and one fourth of firms have not thinking of any metadata strategy itself. Managing metadata through traditional RDBMS is difficult. Now, organizations use enterprise NoSQL technology and semantic data models. Metadata can be structural (e.g., the location of actual data), breakthrough (e.g., author, indexing information etc.), or managerial (e.g., file type, technical details etc). It helps to think of metadata like the glue that harmonizes links and adds context to data. The data that describes about other data is metadata. There are 3 varieties of metadata:

1. Breakthrough metadata – a resource which discovers/identifies elements like author, title, abstract etc.
2. Structural metadata – a resource which should contain data about page ordering for forming chapters.
3. Managerial metadata – the resource that stores metadata about technical details, the creation time, and access information etc.

There exists data of some processes need to be used in some other processes. Even though the data might be similar, but the data structure differs and is unknown when it was built. The elements of multiple processes are ingested in a consistent manner.

II. RELATED WORK

Transportation issues are no way related to document data, separation of functionality is very much required where this is best handled by Envelope pattern [7]. Envelope pattern concentrates on delivery mechanism. Document data should not involve in validation of transport information, security problems, or package issues. The data is separated from transport. The document type which has been created for identifying the domain data and the envelope[7] is used for delivery mechanism. The business data and transport/security information is being separated in envelope pattern. The Data Hub Framework [12] is used for staging documents. It contains the following elements:

- Data creation timestamp
- Source origin
- Batch identifier
- The types of data.
- Data verification.
- Unified resource indicator of document

2.2. Content: This section will have business data unchanged but into valid formats like json, xml etc.Triples[8]: The minute data entity semantics which support data model[1] by providing joins, supports RDF framework[10] Format for retrieving and manipulation of data, Web Ontology language-based reasoning.
Significance of Metadata and Data Modeling of Metadata by using Mark Logic

2.2.1. Harmonized Documents

- The key elements which act as headers would be used for indexing, can make faceting on those elements, and sometimes can be generated as views.
- Date elements can even be used as header elements but should be in compliance with standard date formats.
- The data which is not in original can be stored in header section.
- Original data section – is the section where the given data is stored and will be stored in specific format if required.

Example: {"envelope":{"header":{"firstName":"Lakshmi"},{"lastName":"Prasanna"}, {"pincode":"500072"}}, {"triples":{}}}, {"content":{"businessKey":"value"}, {"businessKey2":"value2"}, {"businessKey3":"value3"}}}

Envelope [7] has multiple sections like header, triples[8], and content sections. The content section can be exposed via REST end points.

We use envelope pattern to store canonical forms of data.

The transformations in Mark logic [10] – Business entities can export from current system, imported to Mark logic as-is, transforming progressively, and relate the documents.

We have verified the metadata data modeling in banking domain and verified the ingestion and search responses, where strategic decisions can be taken at faster pace. Below are the details of the structure being used.

III. EXPERIMENTAL RESULTS

3.1. Indexing structure

Mark Logic uses a distinct parent-child index to identify field hierarchy. The indexing is a fast phrase index where it uses parent-child names as keys.

3.2. Data Ingestion

Data ingestion can be used for ingesting data from multiple departments wherein it can be saved into a common repository. The data may vary, but common fields can be stored in standard repository.

Published By:
Blue Eyes Intelligence Engineering & Sciences Publication

Retrieval Number: B5567128218
3.4. Performance of ingestion and Search Results:
The below figure depicts the performance of ingestion and search results.

![Graph showing Ingestion and Search Performance](image)

**Figure 5. Response Graph Showing Ingestion and Search Performance**

IV. CONCLUSION

In this paper, we tried to understand the importance of metadata and data modeling of metadata by using envelope pattern in Mark logic. We applied these data models in banking domain for different processes like ingestion and search across multiple departments. We have identified, departments can retrieve data and ingest data at a faster pace.

REFERENCES


Authors

Mrs. K Lakshmi Prasanna is a Student of M.Tech CSE from FIT, Chevella, RR Dist – Telangana. She explored in Data Mining, CC,

Dr. J. Sasi Kiran graduated in B.Tech [EIE] from JNTUH Hyd. He received Masters Degree in M.Tech Software Engineering, JNTUH Hyderabad. He received Ph.D degree in Computer Science and system engineering from Andhra Institute of Technology, Chevella, R.R. Dist Telangana State, India. His research interests include Image Processing, Cloud Computing and Network Security. He has published several research papers till now in various National, International Conferences, Proceedings and Journals.

Dr. K. Sreerama murthy graduated in B.Tech [CSIT] from JNTUH Hyd. He received Masters Degree in M.Tech Software Engineering, JNTUH Hyderabad. He received Ph.D degree in Computer Science and system engineering and technology, Yammanpet Hyderabad, R.R. Dist., Telangana State, India. His research interests include Data Mining, IDS, Big-Data analytics , Cloud Computing and Security. He has published several research papers till now in various National, International journals and Conferences, Proceedings .

Published by: Blue Eyes Intelligence Engineering & Sciences Publication