Addressing Data Woes through Effective Metadata Management

Abstract

Compliance activities cost financial services organisations billions of dollars annually and the spending is set to spike on the back of a rise in the number of regulations published by the G20 in recent years. In the absence of automated data management processes, critical to regulatory reporting, the banks are forced to hire hundreds of compliance and risk professionals to fulfil their regulatory obligations across data management, activity monitoring, and transaction reporting.
In addition, external consultants are often employed to implement 'patch' jobs to avoid non-compliance and the consequent regulatory scrutiny and penalties. It is no surprise annual expenditure on regulatory talent and consulting services has been consistently growing.
An effective metadata management platform can go a long way in addressing data issues faced by banks.

This paper analyses the reasons for inefficient metadata management and presents a platform for better metadata management.

The Role of Meta Data Management in Driving Compliance

Audit and traceability of financial transactions is key to ensuring regulatory compliance. In most large banks, financial transactions are spread across multiple applications, databases and platforms. Creating visibility into the sequence in which transactions are processed across these different applications and platforms is effort- and cost-intensive. Metadata holds the information related with different transactions and plays a crucial role in providing insights into the traceability of the data, spanning origination, consumption, reporting and how it is transformed in this journey based on business rules. Being at the core of regulatory reporting, a robust metadata foundation is critical to driving regulatory compliance.

The Root Cause of Metadata Mismanagement

Most metadata management processes - such as, creating inventories of data sources, extracting metadata from them, analysing them, and then generating reports using the metadata - are currently manual and therefore time-consuming and expensive. Additionally, metadata lacks governance because of the evolutionary and temporal nature of applications and technologies associated with IT applications. The situation is further complicated by big data, which brings in new data formats and multiple data sources, both internal and external to the organisation. NoSQL and schema-less databases pose challenges in extracting metadata information from data files (xml or Jason formats), which can have varied formats for the same attribute. Metadata cannot be extracted from such environments without automation and use of parsers and natural language processing (NLP) or machine learning (ML) techniques.

The absence of automated metadata management processes leaves the banks unprepared for compliance with regulations and leads to inaccurate or delayed submissions. This also results in delay in day-to-day impact analysis of changes to data in various systems in production, resulting in data inaccuracy and delays in delivery of changes to systems.

Despite the pressing need to improve metadata management, investment in the area remains low business priority, only driven by regulatory requirements. This need-based approach sometimes costs more. Given the pace of evolution in the regulatory landscape, the financial services organizations need to automate the process of capturing technical metadata in real-time from both static or structured as well as dynamic or unstructured sources.

The Metadata Management Platform

An automated metadata management platform can be compared with a data warehouse in terms of its functionality and usage (see Figure 1). However, it is at a much smaller scale in terms of data volume and usage pattern, and therefore feasible to be created and managed from a build, cost and investment standpoint.

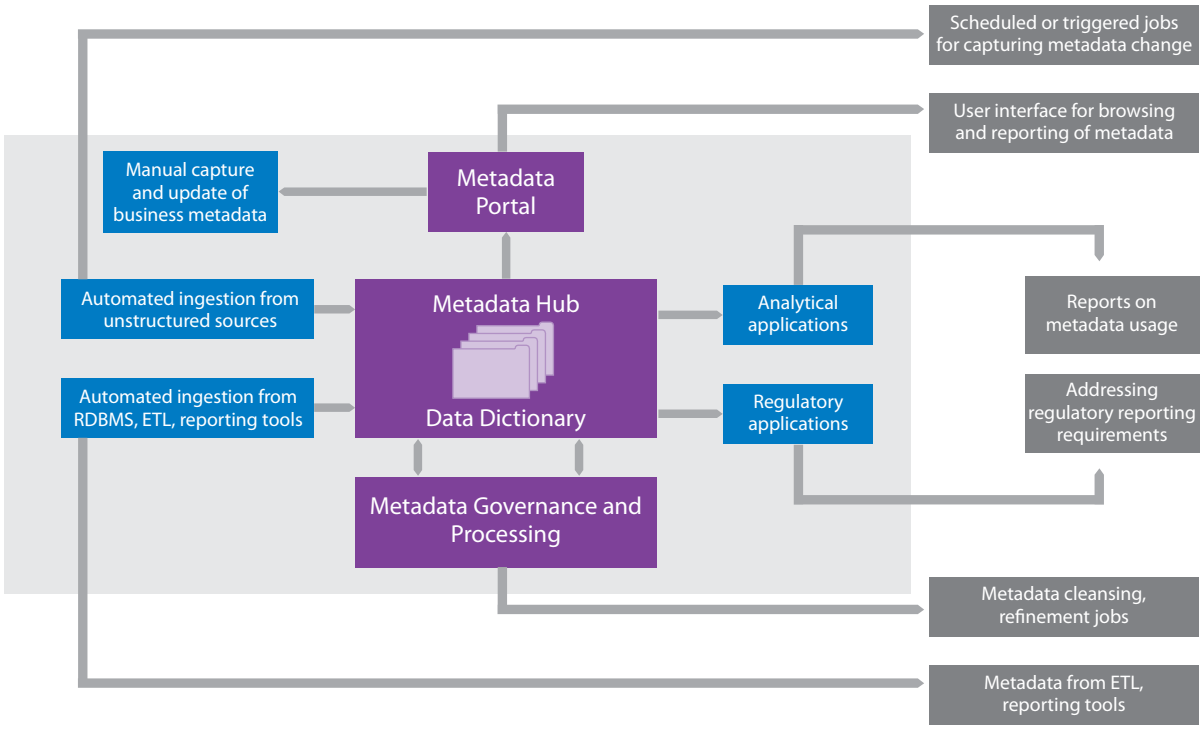

Figure 1: Functionalities of an Effective Metadata Management Platform

Some aspects to be considered while building a metadata management platform include:

Ingestion

An effective metadata management platform will need a process for ingesting metadata information from source systems, which are typically a combination of structured and unstructured application systems. Systems using RDBMS, extract-transform-load (ETL), or reporting tools will be able to provide metadata from their own internal metadata repository and data dictionaries. These are typically technical metadata, however, more critical business metadata - in terms of usage - is typically not stored as part of application systems. Business metadata is either extracted from existing reports by leveraging NLP or parsing through reporting scripts or code or captured through a metadata management portal by a community of business users. Similar to a data warehouse environment, jobs can be scheduled to periodically extract and refresh existing metadata information within the platform. Crawler technologies can be used for extraction purposes.

Processing and Storage

A structured meta-meta model can be used to store the extracted metadata information. Linking technical metadata to business semantics is a key enrichment process that would need a combination of business user inputs and NLP techniques. Identifying contextual usage of data along with meta-tagging can help in the linking process. Metadata information may need cleaning and enrichment for storage and future consumption. Housekeeping jobs can be scheduled for the same.

Consumption

There are several critical use cases for the consumption of metadata information. For instance, the BCBS 239 regulation requires financial institutions to deliver insights into their business data. The General Data Protection Regulation demands data lineage to meet transparency obligations around the usage of data within the organisation. The other important use case is impact analysis for functional changes within applications – a change in an attribute, if not analysed and addressed, causes failures in several downstream applications. Such impact-related information becomes even more critical and costly during merger and de-merger scenarios.

Existing Metadata Management Solutions

Most metadata-management solutions currently use functionality, provided predominantly by extract-transform-load (ETL) tool vendors like Ab Initio (Metadata Hub), Informatica (Metadata Manager) and IBM (InfoSphere Metadata Workbench). ETL tools have to maintain a repository of technical metadata to deliver core ETL functionality. Pure-play metadata management tools, rated by analysts, are also available in the market. The key capabilities of these providers include:

- Easy to use interface for business representatives
- Data inventory with the capability to curate data assets enabled by ML and automatic detection of relationships
- Enrichment, such as tagging, to enable data scientists and stewards to identify and integrate access to additional relevant data assets
- Business semantics, business rules, workflow management, and metadata exchange
- Support for security and privacy requirements

Vendors like ASG Rochade and Collibra NV have been in this space for a long time. There are a number of other product

vendors in the metadata management space. However, each product needs a certain degree of customisation and knowledge transfer to ensure proper use and implementation within a metadata management platform.

Business Participation

There is a general lack of awareness on the importance of metadata management within the business community. Capturing business metadata, linking it to technical metadata and using the information for business purposes will not gain traction in the absence of buy-in from business-user communities. However, the advent of big data, the internet of things and digital applications has attracted attention towards the importance of data. The availability of a portal that can help them understand the depth of enterprise data and its usage can open a plethora of opportunities.

Data lineage capability at the click of a button can increase the credibility of data as well as encourage business users to contribute to the upkeep of the business metadata. This input is critical from a business perspective and it is seldom captured as part of technical metadata. Regular reporting and communication from the platform portal can help infuse a culture of leveraging insights for business operations and decision-making. Understanding the data that can be consumed by data organizations and the need to manage it with care will be key. Needless to say, business users will reap the benefits from the ready availability of metadata information in the form of faster regulatory reporting, analytics, and identification of change impact use cases. Just like reporting portals, metadata management portals will help increase the uptake of metadata management solutions.

Prerequisites to Implementing a Metadata Management Platform

To implement and use the metadata management platform effectively, financial institutions will need to take some preliminary steps:

- Establish an inventory of applications and data stores coupled with network access to all application data stores
- Adopt reporting and ETL tools
- Enable application data stores to access metadata information repositories after considering the security implications related to regular refreshes in an automated environment

- Use crawler technologies to extract metadata
- Get executive sponsorship and buy-in to ensure continuous participation by business stakeholders
- Establish governance standards to monitor participation and usage of the platform

The Bottom Line

Adopting an automated metadata management platform helps eliminate the inefficiencies of manual processes. In addition to reducing the cost of regulatory compliance, an automated platform can help infuse a data-driven culture into financial services organizations. With the two being related, if addressed effectively, this can open up a number of opportunities for financial organisations to unlock exponential value from their data assets – especially given the ready availability of technologies to build such a solution.

About The Author

Bikram Das

Bikram Das is Chief Data Architect, Chief Data Officer Initiatives, with TCS' Banking, Financial Services, and Insurance (BFSI) business unit. With over 25 years of experience in data management and consulting space, he has managed several data initiatives for our global clients across industry verticals. Das specializes in metadata management, data quality, and data governance including the evolving Big Data space. Bikram has a Bachelor's degree in Chemical Engineering from the Indian Institute of Technology, Kharagpur, India.

Contact

Visit the Banking & Financial Services page on www.tcs.com

Email: bfs.marketing@tcs.com

Blog: Bank of the Future

Subscribe to TCS White Papers

TCS.com RSS: http://www.tcs.com/rss_feeds/Pages/feed.aspx?f=w
Feedburner: http://feeds2.feedburner.com/tcswhitepapers

About Tata Consultancy Services Ltd (TCS)

Tata Consultancy Services is an IT services, consulting and business solutions organization that delivers real results to global business, ensuring a level of certainty no other firm can match. TCS offers a consulting-led, integrated portfolio of IT and IT-enabled, infrastructure, engineering and assurance services. This is delivered through its unique Global Network Delivery Model™, recognized as the benchmark of excellence in software development. A part of the Tata Group, India's largest industrial conglomerate, TCS has a global footprint and is listed on the National Stock Exchange and Bombay Stock Exchange in India.

For more information, visit us at www.tcs.com

Experience certainty. IT Services
Business Solutions
Consulting

TCS Design Services I M I 07 I 19