



With **Data as the new oil,**
how to drill for insights

Enabling the decision-makers and their support staff to find trustworthy data is the first step in the insights-to-decisions journey of any organization. However, in today's Volatile-Uncertain-Complex-Ambiguous (VUCA) world, the decision-making fabric of enterprises is in constant flux. Running a successful business has always been based on being able to take the right, fact-based decisions at the right time, and not "gut feel" instinctive decisions. Institutionalizing a data-driven culture is thus a mandate which cannot be ignored any further, especially for breakaway enterprises. But a common problem that we hear from several customers is that while they have implemented many information systems to digitize their business processes and generate tons of data, they are unable to extract business insights and put this data to use.

The importance of being data-driven

To better understand the need for a data driven enterprise, let's consider some interesting facts. **According to an IDC report*, in the year 2019, 40% of digital transformation initiatives will be supported by cognitive intelligence, providing on-time insights for new monetization models. Meanwhile according to Gartner** the year 2020 will see algorithms positively altering the behavior of more than 1 billion global workers.** We could continue to quote several reports and statistics all of which underline the fact that with the advent of digitalization and the growing focus on customer experience, data driven business decisions and operations have become an imperative. Using data and analytics to build an intelligent enterprise is the way to go.

* IDC FutureScape: Worldwide Digital Transformation 2017 Predictions

** Gartner's Top 10 Strategic Predictions for 2017 and Beyond: Surviving the Storm Winds of Digital Disruption

Intelligence of an enterprise is in turn derived from the thorough knowledge of all information assets in possession of the enterprise, and a deep appreciation of all such assets by one and all. That is where an information management software like data catalog plays a big role.

Hallmarks of a typical data catalog product

What is a data catalog? Simply put, a data catalog is a platform to find, curate, collaborate and consume enterprise data and analytics assets. A good one should offer a simple, intuitive and contextual interface for each knowledge-seeking persona. It should enable a faster multi-faceted asset discovery, exploration of disparate assets and seamless integration for persona specific downstream analytical activities.



Hallmarks of a typical data catalog product

In selecting a data catalog platform that can deliver best results, note some key features and capabilities that it should be able to deliver:

- Scan metadata across information assets, including but not limited to structured databases, structured/semi-structured data files, BI/reporting tools, APIs, ML/AI models, unstructured documents and media files, either on-premise or on cloud, and store such metadata in a central repository
- Offer a simple intuitive interface to perform a semantic and faceted search on the metadata and rapidly render search results
- Manage the enterprise level data taxonomy or glossary and make it available to all information seekers as a reference
- Explore each information asset from the search results and analyze the metadata at the asset level. It should be able to show the lineage or relations with other assets if any, refer to associated documentation if any, or raise access request for the asset as applicable
- Drill down and explore the metadata at individual information element that constitute the top level information assets, such as data types, data distributions, variability etc.
- Function as a centralized data governance platform clearly demarcating the bona fide owner of every information asset, with a built-in access request workflow and an approval mechanism to ensure right information is available
- Leverage non-functional features like be hybrid cloud ready, be secure, scalable and fault tolerant and provide high throughput and quick response times

Augmented features of the next gen data catalog product

Apart from the core features mentioned above, a next gen data catalog platform should also offer some advanced augmented features, preferably powered by AI/ ML interventions, as below:

- Auto-tagging, auto-recommendations, auto-relations identification between information assets by leveraging AI/ML algorithms
- Auto-populating and auto-refresh to ensure that the central metadata repository is kept fresh as existing information assets are modified or new information assets are created
- Auto-recognizing "Personally Identifiable Information (PII)" so as to clearly identify sensitive and data elements thereby ensuring data privacy and data encryption
- Enabling data stewards to create federated queries in joining datasets from disparate sources without having to move them from the original location in a self-service mode, and expose such queries as APIs for downstream consumption
- Allowing information consumers to rate the information assets, add written comments, and provide more subjective information about the assets to augment the original metadata
- An intuitive discussion board to obtain and augment more specialized information from an associated SME

How to ensure success with data catalog

While ensuring due diligence in selecting a data catalog product in the first place is vital, its implementation also needs meticulous planning, coordination and communication. Using a data catalog product brings about a change in the mindset of all information consumers within an enterprise. The three pillars of People, Process and Technology must come together to ensure a successful implementation of a data catalog product within an enterprise, large or small. Once successfully implemented, the business and IT community of an enterprise will reap both the short and long run benefits.

Author



Arup Duttaroy

Deputy Head, LTI Mosaic, LTI

Arup heads the Sales Engineering function of LTI's Mosaic platform. He is a veteran with 25+ years of experience in the IT consulting industry, of which 19 have been dedicated to Information Management & Business Analytics. His prior achievements include setting up a 450+ member Analytics Center of Excellence and an Analytics Shared Services team. He holds a Bachelor of Production Engineering Degree from Jadavpur University, Kolkata, India and MBA in Finance & Systems from XLRI, Jamshedpur, India.

LTI (NSE: LTI, BSE: 540005) is a global technology consulting and digital solutions Company helping more than 420 clients succeed in a converging world. With operations in 32 countries, we go the extra mile for our clients and accelerate their digital transformation with LTI's Mosaic platform enabling their mobile, social, analytics, IoT and cloud journeys. Founded in 1997 as a subsidiary of Larsen & Toubro Limited, our unique heritage gives us unrivaled real-world expertise to solve the most complex challenges of enterprises across all industries. Each day, our team of more than 30,000 LTItes enable our clients to improve the effectiveness of their business and technology operations, and deliver value to their customers, employees and shareholders. Find more at www.Ltinfotech.com or follow us at [@LTI_Global](https://twitter.com/LTI_Global)