Enterprise Data Marketplace: Democratizing Data within Organizations

Abstract

Data today has evolved from an asset relevant for reporting, tracking, and auditing, to one that is critical for real-time decision-making. As a result, the harvesting and monetization of data has become an important trend for large organizations, and Gartner predicts that by 2020, 25% of them will either be sellers or buyers of data via formal online data marketplaces¹. Within organizations, this shift in the application of data is driving new approaches to data architectures.

A key impact of this shift is the democratization of data to business employees within the organization. Companies are moving towards self-service analytics and business intelligence (BI) in a bid to encourage the use of relevant, quality data as the basis for better decisionmaking. The business value of this trend has also been established clearly, with Gartner estimating that, "By 2020, organizations that offer users access to a curated catalog of internal and external data will realize twice the business value from analytics investments than those that do not."²

 $\bullet \bullet \bullet \bullet \bullet \bullet \bullet$

.

This paper discusses how the vision of an organizational environment where consistent, quality, and secure data is made available to business users on demand can be achieved. In such a scenario, data can be dragged and dropped from a 'data storefront' to drive meaningful, actionable insights, providing an open self-service analytics infrastructure that follows the economic principle of demand and supply, is available across diverse business functions, and is normalized despite being sourced in different forms and formats (structured and unstructured). The paper discusses the conceptual model of an Enterprise Data Marketplace (EDM) that on the one hand allows business users access to curated and aggregated data as consumers, and on the other, allows them to play the role of producers by contributing to data in the data marketplace.

The Enterprise Data Challenge

In today's world of big data, enterprises have such an abundance of data that they are not able to fully utilize it. They are inundated with numerous sources and formats of data – structured data sets in the form of master data, operational data and transactional data from their customers, partners, employees and vendors, and unstructured and semi-structured data in the form of documents, PDFs, excel sheets, images, transcripts, and so on, about products, regulations, competition, and customers. In addition to the above 'traditional' data sets, there are newer types and sources of data, such as data from internet of things (IoT) sensors, or streaming or crawled web data from internal as well as external sources.

Most data-driven enterprises possess infrastructure and mechanisms such as Extract-Transform-Load (ETL) tools, enterprise data warehouses (EDWs), and data lakes, which help them harness and harvest the data available to them. However, despite these resources, they often fail to leverage the data at their disposal completely or efficiently.



Figure 1: Diversity of Data Types in Organizations

In our experience, there are different levels of maturity of data initiatives in organizations, and the challenges of data harnessing vary across these scenarios:

 Companies with the least maturity of data-focused initiatives are either still struggling with operational issues, or are the ones where, typically, data is held in silos and

 $\bullet \bullet \bullet \bullet \bullet \bullet \bullet$

owned by respective business divisions or departments. This barricades free flow of data and information. At best, these companies have ad hoc ways of sharing data across divisions, and a lot of data stays hidden on individual desktops. This leads to inefficiencies and lack of visibility on important parameters, which subsequently means that inadequate data is available for business SMEs and decision-makers to analyze and draw actionable insights from. Amongst large enterprises, the ratio of such companies is not significant, as they have mostly all evolved to integrated data systems.

- 2. At the next level of maturity are companies that may have an EDW or a data lake initiative, but lack enterprise policies, standards, and (automated) processes (or a combination of these) to get the data into the warehouses or data lakes. The result is missing or inconsistent data in the warehouse, which is not enough to derive real value from. This type of data is also referred to as noisy or dirty data, which requires a lot of effort to clean and make ready-to-use by business analysts and decision-makers. Quite a few large companies are facing such a scenario.
- 3. While a lot of large organizations have overcome the previous two scenarios, they are challenged by the continuously changing environment with new sources of data being discovered and existing sources throwing up new data types and formats. The time they take to adapt with a change in their data management infrastructure is not adequate to protect the business from the impact of inefficient data usage. This typically happens because the data warehouse or data lake was structured and designed with a static view of data based in the past, and catering to the present changes in data sources or formats needs significant effort.
- 4. A few but significant number of large enterprises have been able to upgrade their data warehouses or analytics infrastructures within shortened time cycles to cater to newer sources and types of data. However, even there, the effort is largely driven by the IT department, which brings in some obvious lags and inefficiencies due to the lack of a complete understanding of the business users' needs.

The Enterprise Data Marketplace

Given these challenges, a lot of companies have started data virtualization initiatives to address the ever-changing landscape of data sources and formats in today's dynamic business scenario. These initiatives, however, will fail to address the challenge around IT dependence, as explained in the fourth scenario. The need is to develop self-service capabilities using which the business users and decision-makers can search, discover, and use (analyze/visualize) data by themselves to draw insights. The data provisioned by such a solution would typically constitute all data sets available in the enterprise across internal and external data sources. Further, the curating, cataloging and classification of data from available enterprise data sets would greatly enable the vision of a selfservice-driven and easy-to-use system that streamlines data/information flow across the enterprise.

The following enterprise data marketplace architecture depicts a framework that ensures seamless ingestion, curation, classification, cataloging, and distribution of data. It enables self-help mechanisms for business users to search and discover data, perform analytics, and visualize it in insightful ways.



Figure : Enterprise Data Marketplace Architecture Framework

To leverage the digital wave, therefore, it is important that enterprises prepare a roadmap that helps democratize data within the organization and empowers them to participate in larger data ecosystems.

 $\bullet \bullet \bullet \bullet \bullet \bullet \bullet$

References

- 100 Data and Analytics Predictions Through 2021; Gartner; https://www.gartner.com/doc/3746424?srcId=1-7251599992&cm_sp=swg-_-gi-_dynamic; published June 20, 2017; accessed April 21, 2018
- Magic Quadrant for Business Intelligence and Analytics Platforms; Gartner; https://cdn2.hubspot.net/hubfs/2172371/Q1%202017%20Gartner.pdf?t=1496260
 published February 16, 2017; accessed April 21, 2018

About The Author

Sandeep Saxena

Sandeep Saxena is Program Head in Tata Consultancy Services' (TCS') Research and Innovation unit. His current focus is on developing a data marketplace platform and a cross-domain blockchain platform using existing open frameworks. Sandeep is a TOGAF-certified architect and holds MSc (Hons) and BE (Hons) degrees from BITS Pilani, India.

Contact

Visit the Research and Innovation page on www.tcs.com Email: innovation.info@tcs.com Blog: #Research and Innovation

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 All content / information present here is the exclusive property of Tata Consultancy Services Limited (TCS). The content / information contained here is correct at the time of publishing. No material from here may be copied, modified, reproduced, republished, uploaded, transmitted, posted or distributed in any form without prior written permission from TCS. Unauthorized use of the content / information appearing here may violate copyright, trademark