TATA INSIGHTS AND QUANTS

Impact Quantified

Decoding the Data Mesh
A Practioner's View

March 2023

Introduction

Decentralized Data approach towards organizational decision making

Organizations are investing in a central data team in synergy with domain teams, with the expectation to drive their business based on data. However, the central data team tends to become a bottleneck because they are forced to spend time fixing broken data pipelines and understanding domain data. To address this, the data mesh concept enables domain teams to perform cross-domain data analysis on their own, by providing the necessary data and interconnecting it, analogous to APIs in a microservice architecture. Therefore, this kind of decentralized approach too could help organizations make better and timely decisions based on data.

Data Mesh — a New Concept? or an Amalgamation of existing ideas?

Before we answer the above questions,
Lets understand a Data Mesh in simple terms :

“A Data Mesh is a decentralized data architecture that enables each domain to independently organize and manage their data products” As per Zhamak Dehghani , the creator of the concept, a Data Mesh highlights following principles at its core such as Domain Ownership, Data as a Product, Self-service Data Platform and Federated Governance.

So, is it different from any organization’s data repository that it currently operates in its natural state?

Well, in my view its Yes & No.

For the sake of clarity, we will focus mainly on the aspect of domain ownership and Data as a Product.

Any organization, before transforming themselves to a Centralized Data Lake or Data Warehouse architecture for Enterprise Analytics, will be in its natural state of decentralized data repository;
that is, data is being controlled, organized, and managed by each domain that generates the data.

Now that we have addressed Domain ownership.
Let's see what is "Data as a Product" ?

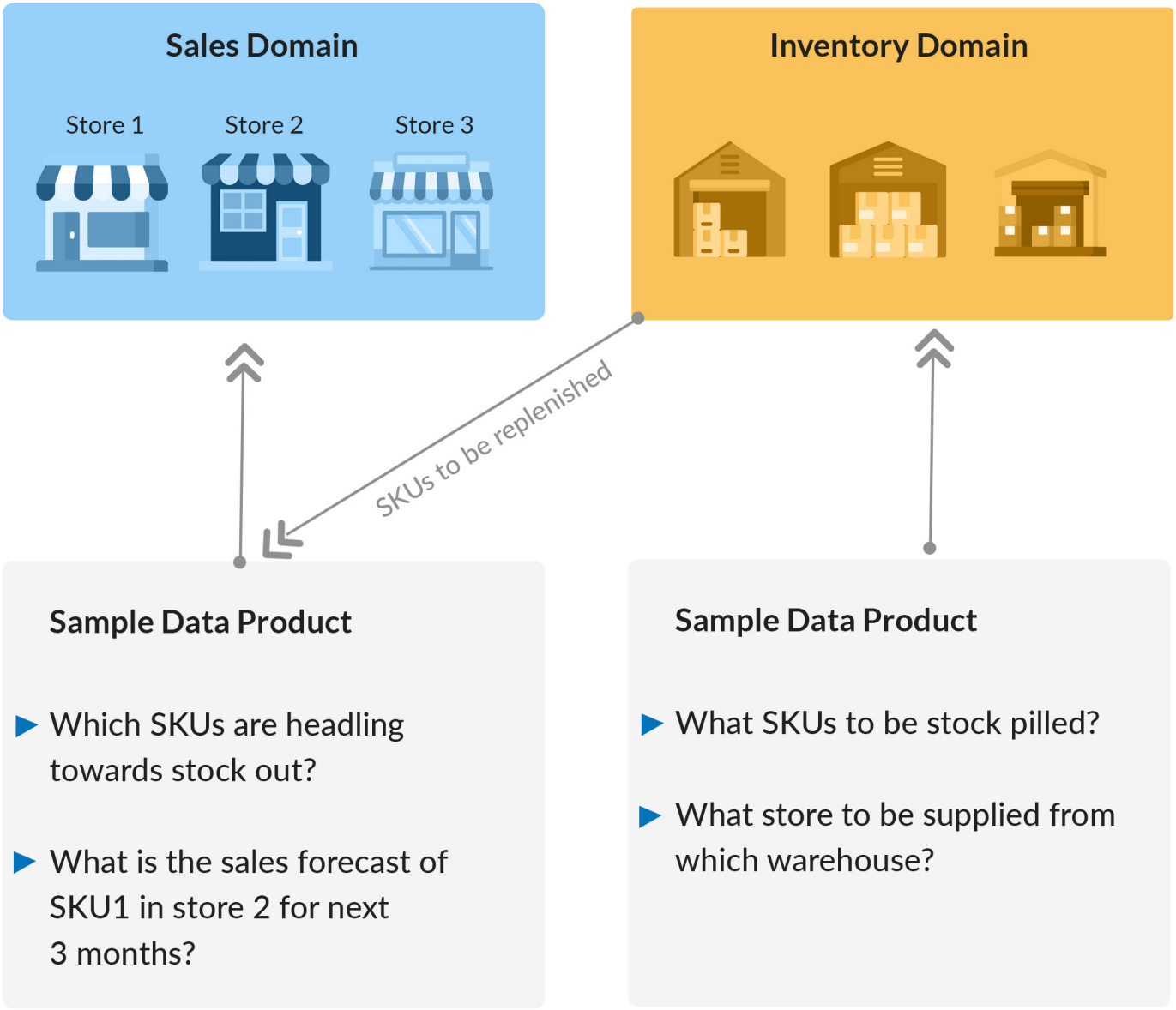
Imagine the Marketing team would like to know how many customers satisfying a particular criterion for a promotion event. Typically, to get this answer we need to aggregate the customer with relevant criteria and then attach it to marketing data. In a Nutshell the No. of Customers becomes a data product in this case an analytical data product.

Data as Product

- How many customers are happy?
- How many customers are eligible?
- What is the best channel for marketing?

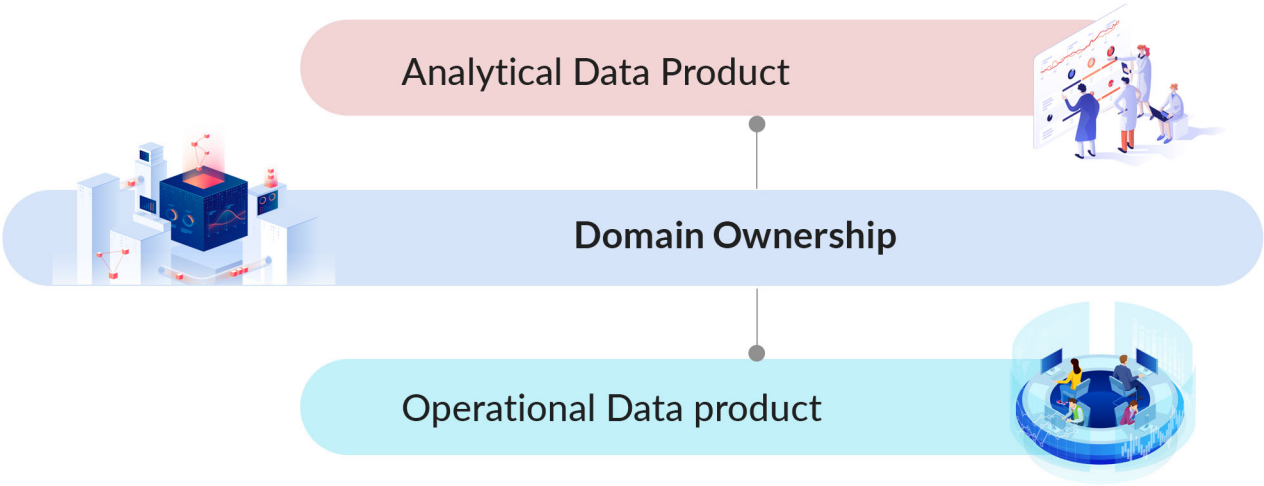
A domain can generate many meaningful data products catering to a particular business scenario and integrate it with various domains.

Let's Understand this with a B2B scenario. A store needs to replenish its stock either from a central or regional warehouse.

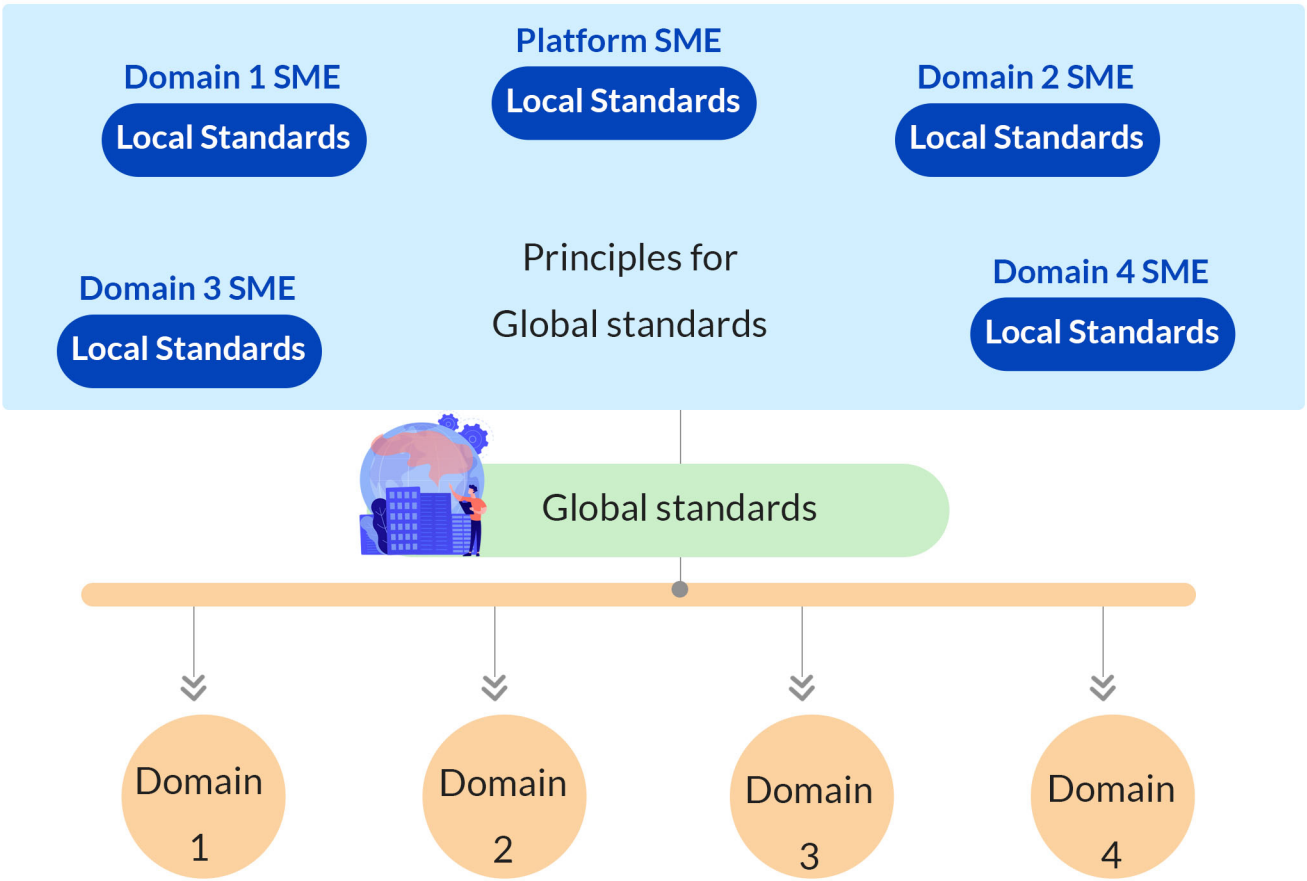


As we notice above the store wise sales as a Domain can create data products like which SKUs are heading for stock out ? Using this Data product, the Inventory management domain can understand the requirement and do the replenishments according to its nearest supply chain which is a data product of inventory management.

The key pattern in this architecture, is that these data products get integrated via a microservice endpoint. This pattern enables a powerful medium of integrating not only the aggregated data but also data from any operational systems with ease.



Enabling this concept via a self-service data platform allows ownership to be retained within the domain and each domain will have governance over the data and the data product they produce.



Federated Governance

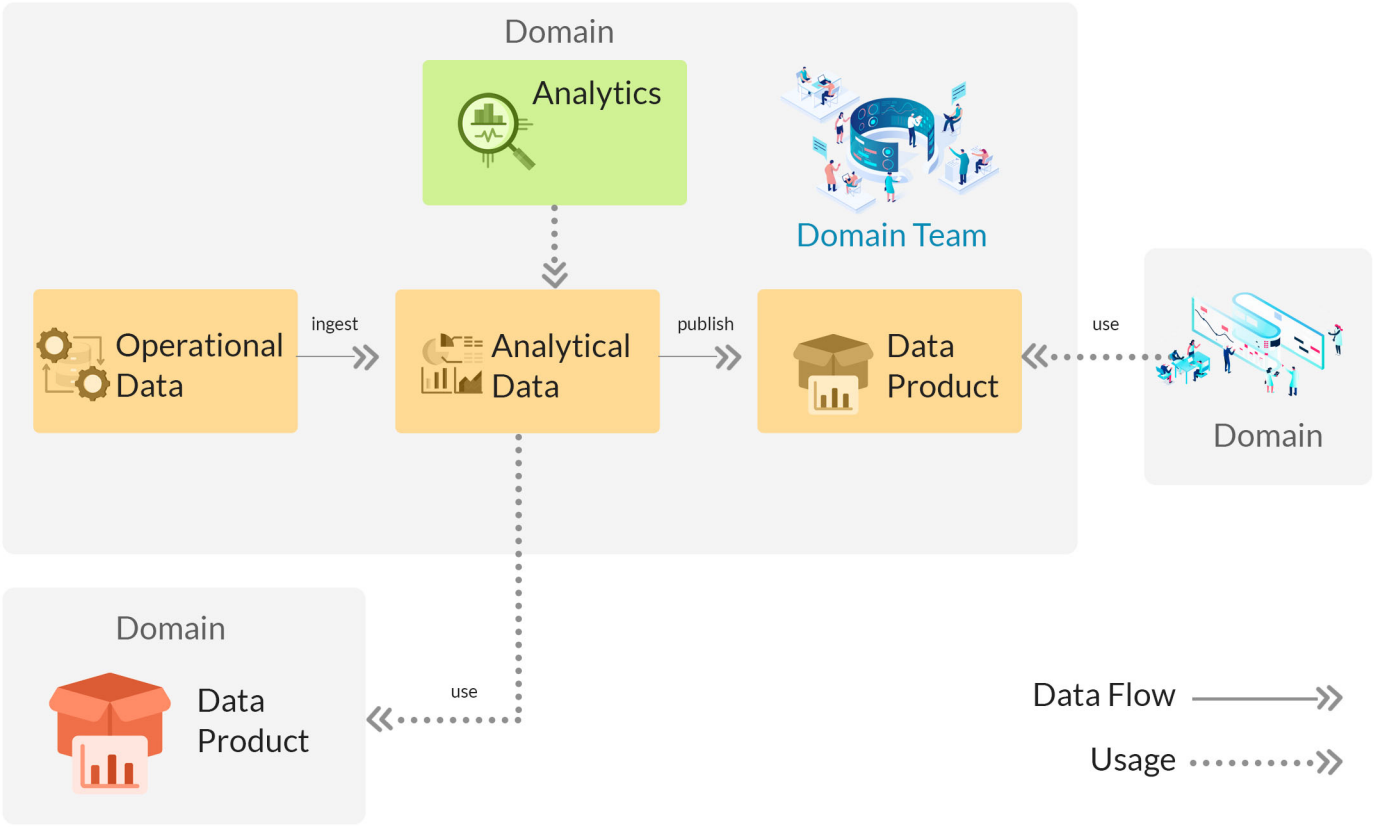
This concept of integrating data via an API has existed since the prevlance of the Service Oriented Architecture (SoA) and Micro Services based Architecture. Now let's answer the question we raised when we began the article.

Data Mesh — a New Concept? or an Amalgamation of existing ideas?

It's obvious that all the individual concepts of Data Mesh such as decentralization of data, Data as product (published via Microservices), self-service platforms, and federated governance are established and long-standing. In our view, Data Mesh architecture is a novel thought of stitching the stand-alone concepts together in a way that enables ease of data discovery and facilitates effective usage of data.

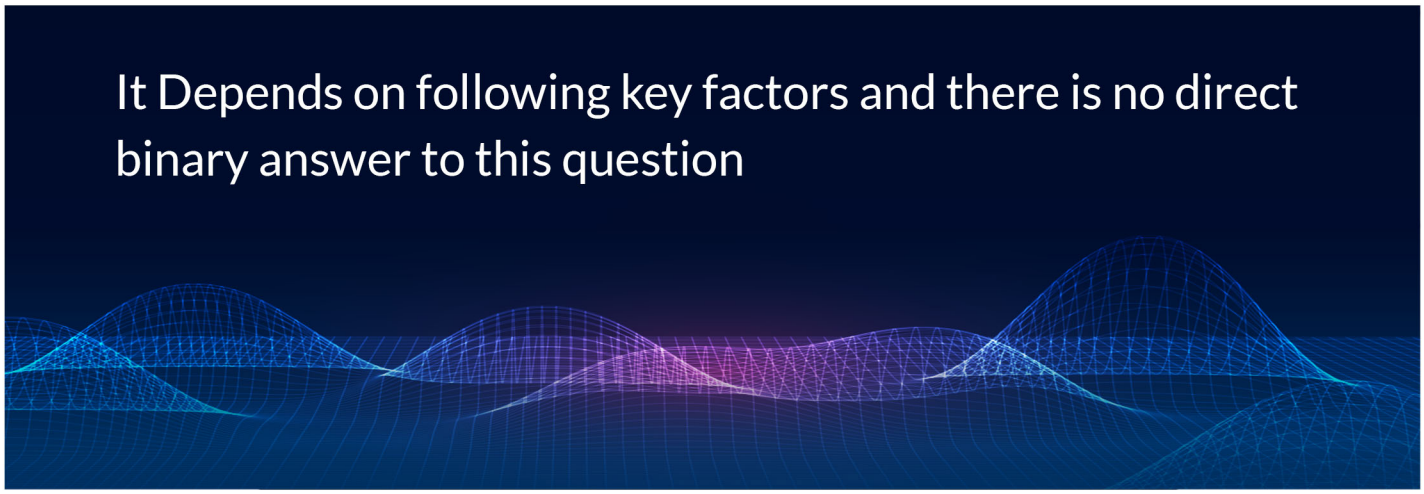
So, how does the data mesh architecture look like?

Let us explore by focusing on a single domain that generates its own data.



The domain's operational system generates the operational data, which then gets aggregated and collated into analytical data.
This domain can fetch data from other domains using the concept of data as the product, and can publish a data product on its own for other domains consumption.

Is a Data Mesh right choice of architecture for every enterprise?



- **Business Problems to be solved:** What are the kinds of business problems in an organization, that needs to to be solved?

- **Appetite for adapting to changes:** How quickly can the organization and its people adapt to technological and operational changes?

- **Number of Domain Teams with Data Skills:** How many functional teams (marketing, sales, operations, etc.) rely on the data to drive decisions by themselves?

- **Number of Domain Teams with Zero Data Skills:** How many functional teams (marketing, sales, operations, etc.) rely on other teams to drive decisions within their own domain?

- **Number of Data Sources:** Total number of source systems that generate data for business consumption.

- **Availability of Data Lake or Data Warehouse:** Whether any data lake or data warehouse has been already implemented?

▸ **Issues in Data Lake or Data Warehouse:** If there is a Data lake or Data Warehouse already implemented, then assess yourself on a scale of 0-10, where 0 means no issues,1 is lowest and 10 is highest, to identify no issues, some issues, or high number or persistence of issues.

▸ **Availability of dedicated Analytics teams:** Is there a dedicated analytics team which helps in leveraging data for decision making?

▸ **Number of people in Analytics Team:** Small (1–5) , Medium (5–8) , Large (> 8) members team

▸ **Data Governance Awareness Level:**

0 — No Awareness
1 — Basic Awareness
2 — Know some concepts and applied
3 — high awareness and rigorously followed

Key factors to consider before adopting the Data Mesh Architecture to avoid future pitfalls

1. Will the analytical workloads deal with massive amounts of data? If so, carefully evaluate the self-service data platform (preferably cloud - based) for consuming the analytical data product from other domains
2. Does the data product consumed using microservice withstand the load?
3. Will there be a need for data processing post consumption of the data products?
4. How is the integrated data going to be consumed for decision making?

References: https://www.datamesh-architecture.com/

Ramji Kannan

Group Leader, Big Data Analytics Solutions

TATA INSIGHTS AND QUANTS

Ramji has over 15 years of experience in the Data & Analytics Industry with experience in the Insurance & Banking domain as a Data and Analytics Consultant and thought leader.

Ramji is currently focussing on building configurable prototypes by applying AI/ML in Data visualization, Simulations, and strategic consulting on Data Office solutions for different Tata Group of Companies.

He brings experience in Machine Learning, Data Mining, Data Management, and Data Visualization techniques.

About Tata iQ

Tata Insights and Quants ("Tata iQ"), incubated as a division of Tata Industries Limited in April 2015, offers multi-sectoral advanced analytics and data engineering solutions using sophisticated predictive analytics and machine learning algorithms.

In our journey since 2015, Tata iQ has added exemplary value to its clients across various sectors. Tata iQ engages with companies, from concept to design, to develop and deliver data solutions leveraging its deep understanding of business and its expertise in data science, to mine meaningful insights for businesses. Tata iQ strives to make a positive impact on the end consumer's life, by delivering faster and better analytical solutions to its clients, which then transcend into faster and better services and products for them.

The specific ideas, concepts, workflows, scenarios, and examples herein are theoretical and hypothetical only, and are intended to illustrate an ideal use-case scenario. A thorough understanding of the organizational infrastructure, intended goals, along with the consultation of an established Data Analytics practitioner, can enable appropriate solutioning which may achieve favourable results. Tata iQ assumes no responsibility or liability in achieving the desired results, nor shall this document constitute any express or implied warranty. Case studies, actual examples, models shared, if any, are property of its respective owners which are shared with Tata iQ as part of a mutual understanding.