

c.Data Architecture and Data Structures

Data Lake and Data Warehouse

Data Warehouse is a centralised repository that contains all the data and handles the works on information. Different with traditional databases is that a Data Warehouse is made for analysis. The data inside is transformed by data mart that gives information about the dimension in analysis.

A Data Lake is a centralised repository that can handles raw, semi-structured or structured data. It can handle large volumes of data and can implement Machine Learning and Data Analytics techniques. It is used as a landing zone, considered a collector of information in which you can take out information.

A Data LakeHouse is a structure that tries to merge the two approach as it is a scalable data architecture that wants to handle structured and unstructured data.

They can be implemented together as data implementation and a staging area. Data transformation at the level you want.

Polyglot Persistence

Way to store the data in different way to achieve different goal. Can use different data storing services to satisfy your needs.

Data Mesh

It is a framework that is a domain driven analytical data architecture framework where data is treated as a product owned by teams that most intimately know and consume the data.

The main principle is to avoid all the mess derived from Data Warehouse/Data Lake. No one know data better than the data owner. There is no multiple departments that decide how to evaluate the data, a single stakeholder responsible of the data and information that the end use give. Data reside by the people in the company that use that data.

Data mesh goals:

- Business: specific business focused on their own data, not ingesting data in a generalised way
- Data Management
- Organization

It is not a centralised model. The data structure focus on centralised domain in the data structure(its focus is not on technologies, it is a framework)

In Data Mesh you are the owner and are in charge of the model with which you handle the data.

Data Governance

You need a data governance strategy to utilise your data to the fullest.

It can be describe as a collection of goal, strategies utilised to speedup the data driven of a company, to increment client satisfaction and compliance and security.

Pillars:

- Data Quality: data clean, accurate, ready to use, fresh
- Data Catalog: data glossary, metadata management, ownership of the data elements, vocabulary of your data, who can see what
- Data Lineage: handles the hierarchy in data, data layers, show how each piece of data is used in the data structure
- Data Privacy: sensitive data should be protected, defines who can access what piece of data, compliance. Data Breach can also derive from an error in the data managing confusing the client to which it is related.
- Data Office: pillar that relies on the base of your organization, is an introduction of data steward(elements whose focus is their own portion of data) and data owner

We have 5 layer that integrates the data in the organization. It is a general flow of data. Data governance can helps to reduce task regarding the addition of data to a structure. Data Governance is also a matter of technology: two technologies used can be Colibra and Informativa. It is an enabler for data as a service and data monetization.

Data as a Service tries to make data and analytics clear and available across all the departments anytime and anywhere. Tries to eliminate redundancy and initiative costs by accomodationg all company data.

Data Monetization is the process that look at company data from a different points of view. Tries to enables some monetary return from the data using a monetization framework.