Data/Information Integration

In current technological era, organizations have multiple data stores and they cater to various types of data which includes structured and unstructured data. Structured data is typically your relational database management systems which includes SQL and No-SQL database. Unstructured data is typically documents, audio, video files, emails, social media data, text messages, chat communications etc. With so many types of structured and unstructured data, integrating or consolidating data together is biggest challenge.

There are ways in which organizations have tackled these challenges first one is through API integration and building web application on top of those API’s. Typically organization are moving to micro services, so each of the system and their sub systems have API’s to access the data in secured manner. With API’s in place it becomes easy to consolidate, visualize and correlate data from disparate systems of records using custom built front end application.

Second approach which is bit traditional, is that organizations have built huge data ware houses. All the structured data from your databases gets pushed to data ware house. Some of the popular data warehousing solutions are AWS redshift, Teradata, Informatica. Organizations use this data ware house to build their business intelligence platform, build reporting, run analytics on top of it and build visualizations. But data ware house solution has one fundamental drawback, it caters to only structured data and does not deal with unstructured data. So to deal with all kinds of data types, lot of big firms are moving towards data lake.

A data lake is a centralized repository that allows you to store all your structured, unstructured, semi-structured data (xml, json) at any scale. You can store your data as-is, without having to first structure the data, and run different types of analytics, from dashboards and visualizations to big data processing, real-time analytics, and machine learning.

In summary data consolidation from disparate systems remains to be a challenge and has been addressed in certain and ways and in future there could more alternatives as well.