Megan Moore
Buff ID: 1014735

Getting a consistent view of business performance is tricky. This involves comprehensive data integration from all data systems within a business. Data integration is the process of using multiple sources and combining them into a single, unified view. The integration process of data begins with data ingestion followed by data cleansing, mapping and transformation. This allows and end result of data integration to enable usage of analytical tools for producing efficient and actionable business intelligence. Integration helps business reach success through benefits such as improved collaboration among teams, boosted efficiency from more valuable data and time savings through reduction of rework. For data integration to be successful, it is important for understand what data is needed by the end user and where that data is stored. Without unified data, one single report could involve gathering multiple sets of data across multiple accounts along with repeated copying, reformatting and cleansing. No analysis could even begin until those steps are taken. Analytics need to be current and efficient. Therefore, being able to conduct all the data operations as efficiently as possible illustrates the importance and necessity of data integration.

There is no one size fits all approach to data integration. The process varies by each business need and structure. ETL, also known as extract, transform and load, is a subprocess within data integration. This is when data is taken from the source system and put in to the data warehouse. This is ongoing within the data warehouse so multiple data sources can be combined into useful information for widescale business intelligence and analytical efforts. Data integration and ETL are two important concepts in the field of data analysis. Both collect data from multiple sources for analytics and use. However, not all data integration solutions use ETL specific tools or procedures. Where they vary is in terms of scope, output, data volume and tools used. Data integration can involve in size of data whereas ETL is most typically used for large

volumes of data. Data integration encompasses a wide scope of activities as listed previously from data ingestions, cleansing, transformation and finally distribution. ETL is a more narrowed approach focused on extraction, transformation and loading of data. The output from ETL is usually a destination system, such as a data warehouse whereas the output of data integration can be one single organized view for reporting ready for a specific application.

In many cases, businesses lack one single source of data and this makes it challenging to answer questions that arise when leadership is seeking to review their strategic objectives. For data integration to work well, specifically the ETL process, business rules must be defined which explain how the data should be routinely transformed. Without this groundwork, data integration overall will not produce helpful results or in the worst case, fail. There are numerous use cases for data integration ranging from reporting, analysis, decision making, data warehousing and business intelligence. Examples of industries using data integration are far reaching from retail to healthcare, finance and manufacturing. Some common uses for ETL are data warehousing, data migration and real time data processing. In the context of data integration, ETL has a crucial part to play for enabling organizations to create a comprehensive central repository pulled from any number of sources. The ETL process helps bring the data together in a consistent and cohesive way. Both tools are useful in managing and manipulating data even though they have different purposes.

The amount of unstructured data in companies continues to grow every day. The need for data integration will always exist for improving decision making and customer experiences while also streamlining organization’s operations and productivity. Data integration overall must keep evolving to keep pace with the ever-changing needs of organizations, leaders and the types of data being collected.