Semi-Structured and Unstructured data

Nowadays, there is a tremendous amount of data in the world and we are still adding more data with an approximate speed of 2.5 quintillion megabytes daily (Bulao, 2022). According to MongoDB, about 80 to 90 percent of the data that organizations collected are semi-structured and unstructured, and the value of the data is enormous. While structured data are specific and easy to process, unstructured data provides more native and competitive insights into the data. Southwest Airlines is the company that we looked into for the project and we thought that graphs and image & audio are the two major semi-structured and unstructured data that can bring massive value to the company.

The graph is a very classic unstructured data that represents the entity as a node and the relationship of the entities as an edge. There are many benefits for Southwest Airlines to use graph-structured data. One of the most commonly used techniques is the graph-based price engine which gives the airline company to dynamic pricing (Peart, 2020). This allows the airline company to offer its customers any price at any time on continuously changing fare prices, which was not as easy before the graph was used because of the complexity of adjusting fare prices among destinations. Southwest can use the same technique to offer its customers the lowest price ticket or the shortest flight time ticket. Southwest can also use it to eliminate or merge flight routes that have a significantly small amount of customers and unprofitable revenue. The way we want to integrate this with the OLTP system is that we will have a graph database system, like OLAP, connect with the OLTP. We can send data from multiple OLTP to OLAP to do graph analysis.

Second, we suggest Southwest Airlines use images and audio. The two main images we are going to use are damaged-luggage images and airplane part images. With the images of the damaged luggage, Southwest can investigate the reasons for the damage and avoid the damage if they find out that the damage is caused by the transportation team during the loading

and unloading phase. In this case, they will provide better training for the transportation team. In addition, Southwest should use images for airplane part defection detection since detection is more efficient and accurate by using machine learning techniques. Audio is also a valuable data source. Southwest can use the flight attendance recording to give better training lessons to the new employees and make a better pre-recoding announcement. They should also store the record of the black box (Flight recorder) when there is an accident. The company can use the recording to investigate the reason for the accident and avoid the same thing happening again. Both images and audio can be stored in the cloud, like an AWS S3 bucket, and we can store the result of the analysis in the maintenance OLTP system and the baggage OLTP system. This can improve the customer experience of the flying process and the safety of the flight.

To sum up, Unstructured data is very important and useful, and many other unstructured data can be used to generate business value such as comments about the flight, we can use it to do sentiment analysis to make a better service for increasing customer satisfaction. Southwest Airlines should invest in exploring unstructured data and design solutions that can integrate unstructured data with its database.

Citation

- Bulao, Jacquelyn. “How Much Data Is Created Every Day in 2022?” *Techjury*, 26 Nov. 2022, https://techjury.net/blog/how-much-data-is-created-every-day/#gref.
- Peart, David. “The World's Airlines Rely on Graphs to Offer Travel Options for Customers.” *Enterprise Times*, 26 Feb. 2020, https://www.enterprisetimes.co.uk/2020/02/27/the-worlds-airlines-rely-on-graphs-to-offer-travel-options-for-customers/.