WQD7007 Big Data Management Assignment:

AliExpress

Sritharan Sivaguru (17198431/1)

Kuan Nam Sung (17029014/1)

Liu Hong Yang (17201091/1)

University of Malaya

Table of Contents

Summary

The company we've identified for this assignment is AliExpress, an online retail provider that operates under the Alibaba Group which is the largest online retailer in China. While Alibaba is itself an online retail provider that focuses on wholesale business and manufacturers, AliExpress focuses on online retail that is much smaller in quantity from smaller businesses or suppliers as shown in Figure 1. It provides international online shopping services around the world and builds a cross-border platform to connect sellers and consumers over the Internet (Alibaba, 2010). As its number of users on their platform has soared dramatically in recent years, the big data technologies that support the transaction and security of AliExpress are also evolving to meet the companies' demands. In this report, we investigate and introduce the big data characteristics and technology used in AliExpress or Alibaba group to generate insights from its treasure trove of e-commerce data.

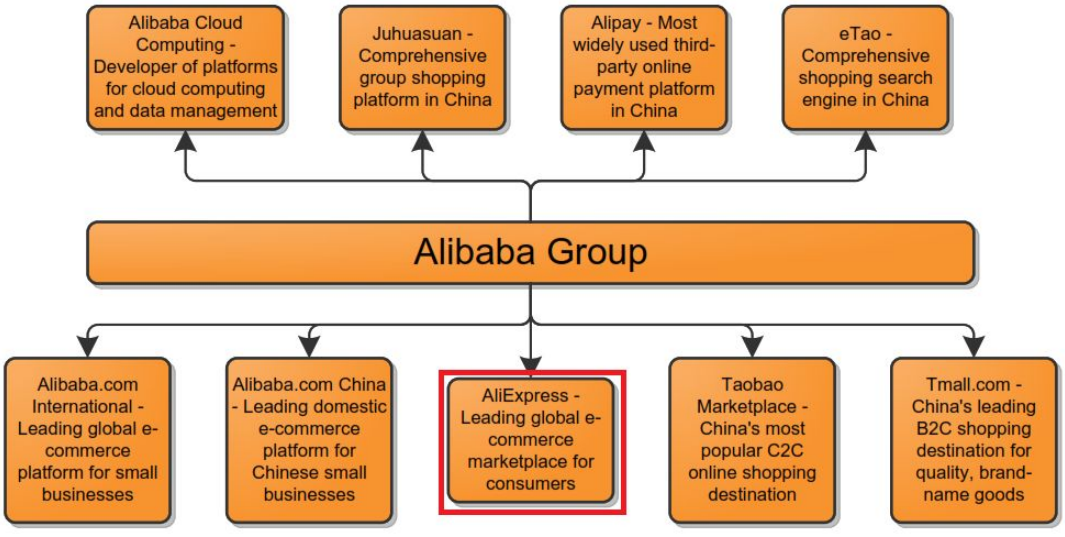

Figure 1: AliExpress within the Alibaba Group (Gong & Kan, 2013)

AliExpress fitting into the characteristics of Big Data

There is an ever increasing demand for cross-border trade from manufacturers and suppliers that are based in China, especially with online e-commerce that offers an online retail experience. It is expected that the global cross border e-commerce trade would grow as much as US$ 1 trillion by this year, exceeding 943 million e-commerce consumers worldwide (Zhan et al, 2020). In this section of the report, we explore the affinity of AliExpress for big data from the aspect of data volume, variety, velocity (speed) and veracity (accuracy). We would also explore the opportunities for visualization and value.

Volume

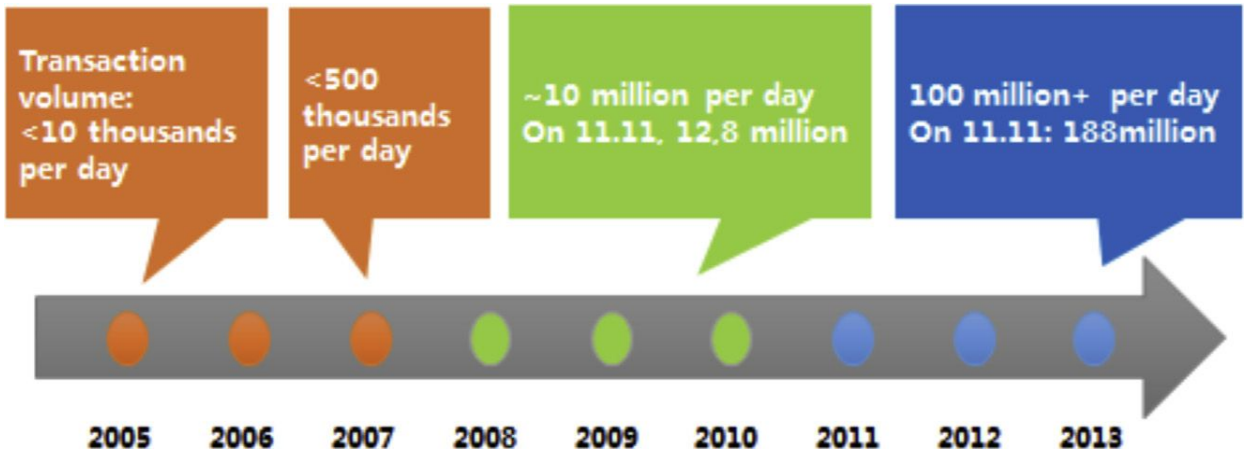

Figure 2: Transaction volume trends at Alibaba (Chen et al, 2015)

AliExpress being China's largest cross-border e-commerce retail platform, covers 230 countries with transactions exceeding 1.5 billion (Zhan et al, 2020). Due to the sheer market reach and consumer size, it is expected that the volume of data would be massive. Even though there is no available indication of the size of each transaction, our research uncovers that the volume of

transacted data on the Alibaba platform utilized by AliExpress, was previously reported at 100 million transactions per day as shown in Figure 2. It could be speculated that the approximately 10EB data store (Chen et al, 2015) has continued to climb since.

Besides the transaction volume of consumer purchase data, the data volume would also extend to the supply chain to track and deliver the purchased goods. It is important to note that not all the data may reside with the e-commerce platform, as the data interchange between suppliers and AliExpress would also generate massive volumes of data for end to end purchase and fulfilment (Gong & Kan, 2013).

Variety

To complete the purchase and fulfilment of transactions, AliExpress would also need to deal with both structured and unstructured data. The data stores would be made up of structured data records, unstructured text, images and documents. This would be required to address presentation of item listing to consumers for purchases, right up to customs clearances for item delivery (Gong & Kan, 2013):

- Purchasing: structured data generated through supply processes, back-end sourcing and ordering, front-end payment of products;
- Warehousing: structured data generated through storage and handling of products in warehouses;
- Delivery: unstructured data generated to physically move the purchases to consumers;

- Sales: all the structured and unstructured data generated through the interfaces that deal with customer demand, e.g. item listing/description and pricing, stock availability indication or forecasting, order processing.

Velocity

Customers typically expect their items to arrive as soon as possible upon ordering and payment. While there are time and physical barriers to AliExpress' delivery fulfilment, the swift processing of orders is a common expectation of e-commerce consumers. Customers need immediate confirmation that their orders have been secured upon payment and that they only need to wait for it to arrive. These key expectations provide data metrics for service quality (Bai et al, 2018). Then there is the need to provide and swiftly process both historical and real time data to provide crucial analytics of consumer product information consumption, search behaviours and eventual purchases (Zhan et al, 2020).

Veracity

The accuracy of the data gathered from consumers, sellers and suppliers with minimal loss as much as possible, is crucial in personalized consumer targeting, application of dynamic pricing methods, customer service quality improvements, analytical forecasting of consumers and goods, and a transparent supply chain for effective procurement/logistics. At any point the veracity of the data gathered fails, effective decision making for the demand and supply would also fail (Lukicheva & Semenovich, 2018).

Variability

Data on AliExpress could vary from pre canned data with selectable options to data that could greatly vary from one item listing to another. Besides the variations in item descriptions, there exists the consumer comment/review system with varying response structures and languages. In this case, buyers could write free form text describing their shopping experiences in their native language and rate their sellers. These variable data sources could be analysed using statistical methods to determine buyer concerns and their nuances (Zhang & Zhang, 2012).

Visualization

Due to the AliExpress platform recording data about their users and sellers on their platform, a massive amount of data could be analysed about their behaviour. Correlation analysis, regression analysis and cluster analysis could be performed to visualize customer behaviours (Wu & Yao, 2017). An example of the analysis and visualization that could be performed is shown in Figure 3.

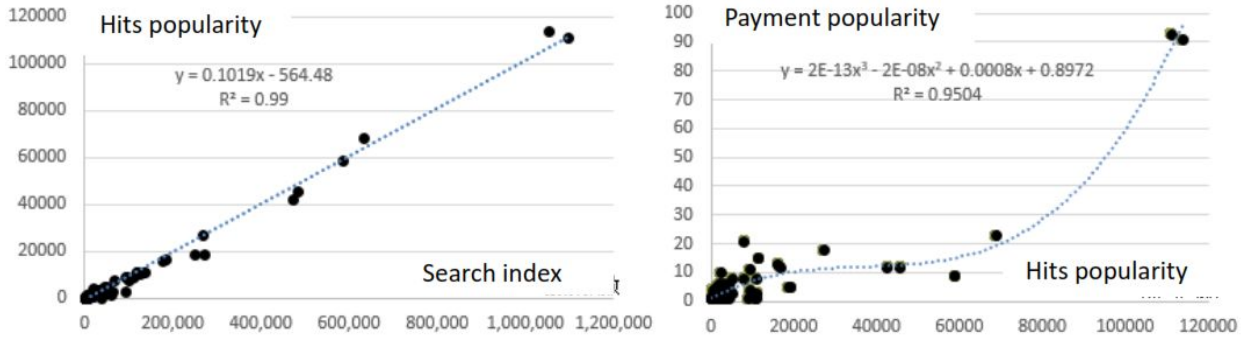

Figure 3: Sample regression analysis of AliExpress users (Wu & Yao, 2017)

Value

The success of any e-commerce platform is dependent on the efficiency as well as the effectiveness of its distribution networks and fulfillment strategy (Gong & Kan, 2013). Having the ability to analyze and derive insights from its wealth of consumer and supply chain data is crucial to its survival and competitiveness. This is achievable through the use of big data solutions and strategies..

AliExpress' dependence on Big Data solutions

With AliExpres being part of the Alibaba group, it benefits from leveraging the shared platform and technologies that power the e-commerce services in the stable, namely Taobao and Tmall in China. The transaction volume shown in Figure 2 illustrates the potential for Alibaba's business growth and unavoidable data expansion.

While researchers within Alibaba have stated that the group optimizes its big data processing capability constantly to improve it's processing prowess for large amounts of data, the group's data platform had its humble beginnings in solutions from Oracle in 2009. Eventually, the group migrated to ODPS in 2015, achieving near real time data processing capabilities. The newer ODPS solution (recently rebranded as MaxCompute) is able to offer a data warehouse platform to manage a massive and varied spectrum of data in a distributed way. Furthermore, the new data processing solution assists the group to address the roadblocks in managing data volume and data streaming effectively, which is near impossible to do now using the antiquated Oracle data

platform.

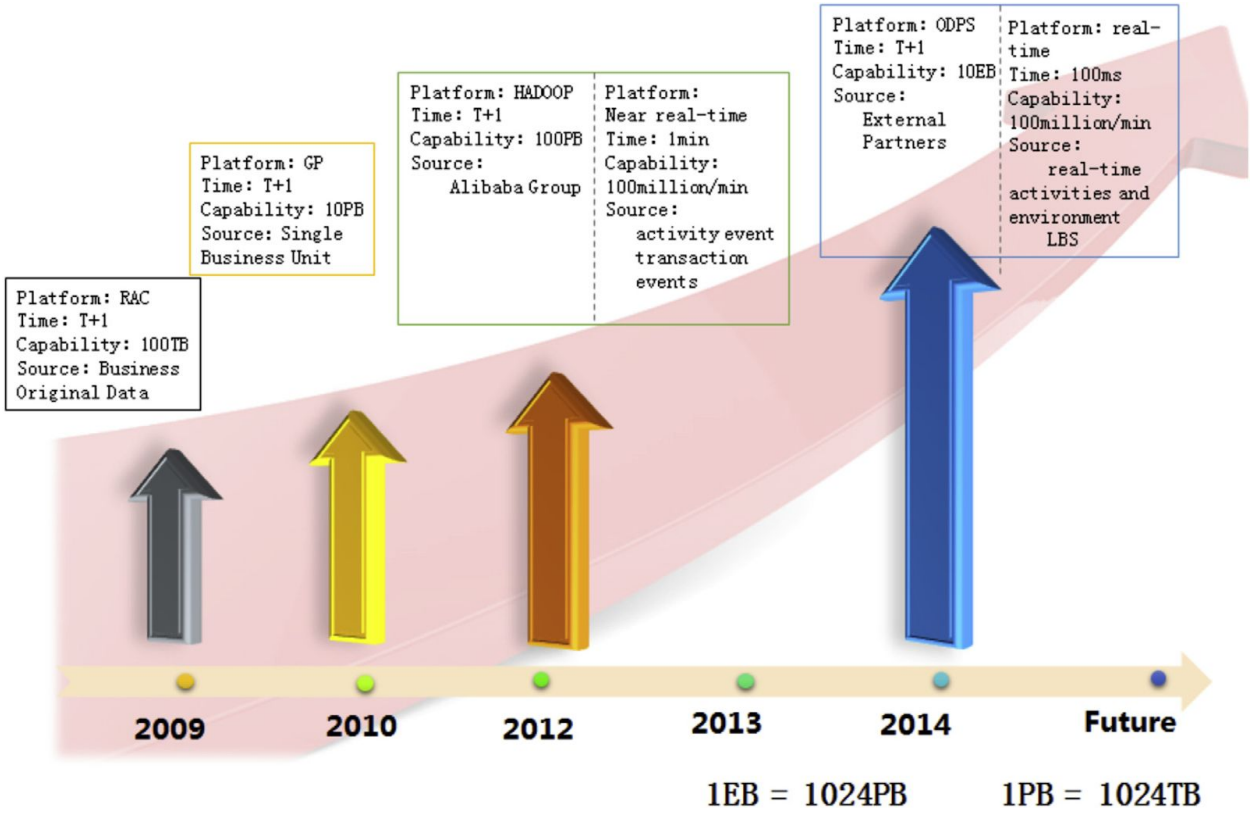

Figure 4: Big data computing process (Chen et al, 2015)

Figure 4 illustrates the journey to the ODPS big data solution, that has assisted the group in solving their big data issues. The first problem was with the data storage capability, which increased requirements from 100TB in 2009 to 10EB in 2014. With the new solution to handle various data sources, data processing time had also decreased dramatically. We've found that the new generation processing system boosts the data storage and data analysis performance of massive data (Chen et al,2015).

An obstacle solved by Big Data: The recommender system in Aliexpress

With the extraordinary performance improvements on big data technology within the Alibaba group, AliExpress is capable of running machine learning on the ODPS big data processing system. The company collects large amounts of personal data from registered customers to build a recommendation system and uses it to sell online products based on these users' preferences. The two authors Lukucheva and Semenovich (2019) conducted a survey of the big data technology used in the AliExpress recommendation system to determine their advantages in enhancing the loyalty of Russian customers buying goods from online retailers. The survey shows that in order to build a recommendation system, customers on AliExpress are required to provide their data and collect these data during the usage of AliExpress. The platform utilizes big data tools to analyze the user's personal information, such as likes, comments, or browsing information, etc. It then executes the recommendation algorithm to predict the user's interests or preferences on similar goods. In this case, big data tools could be used to also help customers personally find what they like, and provide them with better services when using online shopping sites. Furthermore, after utilising this big data analyzing technology, the revenue of Aliexpress in the Russian e-commerce market had continued to increase since 2010 (Zhou et al, 2017).

Contributions

- How the identified company/institution fits into the characteristics of the Big Data, including Volume, Variety, Velocity, Veracity, Variability, Visualization and Value.
    - AliExpress fitting into the characteristics of Big Data (Sritharan Sivaguru - 17198431/1)
- Discuss why the company/institution needs the big data software.
    - AliExpress dependence on Big Data solutions (Liu Hong Yang - 17201091/1)
- What are the key obstacles faced by the company/institution to implement big data solution effectively
    - An obstacle solved by Big Data: The recommender system in Aliexpress (Kuan Nam Sung - 17029014/1)

References

Zhan, Z., Li, Z., & Guo, M. (2020, March). Research on Data Selection of Cross-Border Retail E-Commerce Enterprises from the Perspective of Consumer Search Behavior—Take AliExpress, a Cross-Border E-Commerce Platform, as an Example. In *5th International Conference on Financial Innovation and Economic Development (ICFIED 2020)* (pp. 378-387). Atlantis Press.

Gong, W., & Kan, X. (2013). Logistic Service Development of E-commerce: A case study of AliExpress-an online international trade platform in China.

Bai, J., Chen, M. X., & Xu, D. (2018). Search and Information Frictions on Global E-Commerce Platforms: Evidence from Aliexpress. *Available at SSRN 3267094*.

Lukicheva, T., & Semenovich, N. (2019, November). Big Data as a success factor of AliExpress in the Russian market: advantages and opportunities as seen by the eyes of consumers. In *Third International Economic Symposium (IES 2018)*. Atlantis Press.

Zhang, X., & Zhang, N. (2012). What dose Online Wholesalers Care? Analysis of Text Comments Based on Aliexpress Data. In *WHICEB* (p. 47).

Wu, Y., & Yao, L. (2017, May). Cross-border Consumer Research Based on Cluster Analysis of Aliexpress Behavioral Data. In *2017 International Conference on Applied Mathematics, Modelling and Statistics Application (AMMSA 2017)*. Atlantis Press.

Chen, J., Tao, Y., Wang, H., & Chen, T. (2015). Big data based fraud risk management at Alibaba. *The Journal of Finance and Data Science*, *1*(1), 1-10.

Zhou, J., Li, X., Zhao, P., Chen, C., Li, L., Yang, X., ... & Qi, Y. A. (2017, August). Kunpeng: Parameter server based distributed learning systems and its applications in alibaba and ant financial. In *Proceedings*

of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 1693-1702).

Online Resources

Alibaba.com, (2010, April 26). *Alibaba.com Launches AliExpress to Bring Industry Leading Security, Choice, Flexibility, Convenience and Profitability to Small Wholesalers and Retailers*. Retrieved from https://www.alibabagroup.com/en/news/press_pdf/p100426.pdf