Big Data Case Study

Table of Contents

Organisational Overview

The organization is an American multinational brand that serves customers with different products available in the market. It includes the e-commerce business that the organization has started in the year 1994, by Jeff Bezos. The founder started selling books online on the internet from his garage. The organization currently is one of the biggest e-commerce brands. It allows the customers to shop the products across the globe and get the products delivered at their doorsteps. The role of the organisation is to provide the customers with the delivery of different products anywhere like home appliances, books, and other products. It collected the data of the users based on which the shipping, tracking, marketing and other operations of the organisation take place.

Need

Big data is the big volume of raw information of the customers that can be extracted with the help of accurate analysis. It includes the focus on the generation of data at a very large quantity and high speed every day. Businesses utilise big data for commercial purposes to serve their customers. The organisation needs big data as it allows the organisation to operate with their shipping model based on big data. It allows Amazon to predict the products that the customers are most likely to buy. It allows the organisation to analyse the purchase pattern and market the products and services. The price optimisation is also facilitated, and thus the below information focuses on the use of big data in amazon, the challenges faced by the organisation and how the organisation overcome those challenges. The time required for the completion of the project (Oussous et al., 2018).

Challenges

The use of amazon's product recommendation is most familiar in the application of big data and it includes the collection of insights. It allows the organisation to provide the products and recommendations based on things that are already in carts and the history of orders. Organisation uses the Amazon Aurora, Athena, redshift, S3 and S3 analytics, and Apache Spark, etc. Organisation currently has 1,000,000,000 GB of data that uses more than 1,400,000 servers. Challenges include data growth issues, securing of data that increased the data leakage and

breach risks for Amazon. The data breaches issues are required to be considered like the issues related to the public profile on Amazon, recordings of Alexa, public product list, browsing history of the customers, Alexa data sharing, etc. are the issues required to be considered for the data breaches and security purposes (Wang et al., 2020).

The technical challenges include the unavailability of data science professionals, validation of data and accumulating data from various sources by Amazon. The data set required was approx. 1,000,000,000 GB than requires the availability of 1,400,000 servers. The data accumulating id fast as it includes the focus towards the collection of data when the users are using the website, at the time of purchase and shipping address details, etc. Amazon S3 includes the CSV or delimited text files, the data format is a string data type with a data format that cannot be changed, it is in the String 256 format that is hard to format.

Stakeholders

The stakeholders include the employees, customers and the top-level management. The customers and employees are the key stakeholders as their personal information is stored and it includes the employees of the organization who are required to be updated with the sensitive details. It includes the focus towards society as it is also the social responsibility of the organization to protect the user data. It includes the fulfilment of the customer requirements related to the enhanced services and the protection of confidential data. The recommendations and different information or marketing are done by the organisation make it easier for the prospects to find the products required (Woo and Mishra, 2021).

Requirements, resources needed

The resources utilised include the high-end professional devices, servers and the use of software like S3 and AWS, etc. The resources were both mixed i.e., commercial and also internally developed which allows to balance the risk and reduce the dependency on the human resource. The organisation requires professionals' developers, coders, and professionals who can manage the networking and other tasks linked with the website. It also requires the lead generator or business development manager, most importantly the research and data analysts. It includes the use of the high-end gadgets, AWS, and the use of 4G network that is required to be considered

and it includes the use of Amazon EMR that facilitates the analysing, processing and applying ML to big data utilising open sources of connection (Chong et al., 2017).

Time

The time required for the completion of the project was 1 year and by the end of 2020 that includes the focus on 12 months in which the organisation will focus on the time enhancement of the big data collection and overcoming the challenges of big data.

Results/Findings

Results include that the big data through the use of S3 and EMR is stored effectively though there are privacy concerns for which Amazon is required to enhance the focus towards the security of user data. The project was successful, and the surprise discovered includes the focus on the amount of data the organisation is storing. The lessons learned during the project include the essential need for high-end software, and cloud computing tools for big data storage that ensures effective and use and better security of data. The actions taken include the enhancement of the security to prevent data breaches that ensures the enhanced level of value to the stakeholders (Hariri et al., 2019).

Critique

The project could have been improved through the enhanced utilization of the technologically advanced data collection tools and the upgraded use of could base storage that supports unlimited storage. It includes the major focus on the improvement through the recruitment of good research and data analysts.

Explain

Amazon AWS

It is the cloud-based Amazon web service that is utilised for sharing and storing the data for the businesses and ensuring scalability.

Amazon S3

The S3 means a simple storage service that allows the scalability, security and availability of data to the organisation.

Amazon EMR

The term refers to Elastic MapReduce that is a web service allowing quick and easy processing of vast amounts of data, and cost-effective method used by Amazon (Zhu et al., 2018).

References

Chong, A.Y.L., Ch'ng, E., Liu, M.J. and Li, B., 2017. Predicting consumer product demands via Big Data: the roles of online promotional marketing and online reviews. *International Journal of Production Research*, *55*(17), pp.5142-5156.

Hariri, R.H., Fredericks, E.M. and Bowers, K.M., 2019. Uncertainty in big data analytics: survey, opportunities, and challenges. *Journal of Big Data*, *6*(1), pp.1-16.

Oussous, A., Benjelloun, F.Z., Lahcen, A.A. and Belfkih, S., 2018. Big Data technologies: A survey. *Journal of King Saud University-Computer and Information Sciences*, *30*(4), pp.431-448.

Wang, J., Yang, Y., Wang, T., Sherratt, R.S. and Zhang, J., 2020. Big data service architecture: a survey. *Journal of Internet Technology*, *21*(2), pp.393-405.

Woo, J. and Mishra, M., 2021. Predicting the ratings of Amazon products using Big Data. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *11*(3), p.e1400.

Zhu, L., Yu, F.R., Wang, Y., Ning, B. and Tang, T., 2018. Big data analytics in intelligent transportation systems: A survey. *IEEE Transactions on Intelligent Transportation Systems*, *20*(1), pp.383-398.