  **ITRI623: DATABASES**

**ASSIGNMENT 3:**

**DATA MIGRATIONS**

**8 September 2022**

**Llewellyn Anthony 32969694**

**Prof. R. Goede**

**TABLE OF CONTENTS**

[**MOTIVATION FOR CHANGE** 1](#_Toc112656040)

[**CHALLENGES AND SOLUTIONS** 1](#_Toc112656041)

[**DATA TYPES** 1](#_Toc112656042)

[**DATA MIGRATION TOOLS AND SCHEDULES** 2](#_Toc112656043)

[**CONCLUSION** 2](#_Toc112656044)

[**BIBLIOGRAPHY** 2](#_Toc112656045)

**LIST OF FIGURES**

No table of figures entries found.

**LIST OF TABLES**

No table of figures entries found.

# **MOTIVATION FOR CHANGE**

Walmart is renowned for using data to alter its business strategies and creating software that can monitor customer behaviour in real-time from bar codes scanned at its checkout registers (Jha *et al.*, 2014:9). Big Data analytics were employed to predict market demands, offer timely replies, and boost corporate performance (Chan, 2020:20). Walmart invested properly in its IoT technologies via using biometric sensors in shopping carts, among others (Chan, 2020:20). Social media was used to attract consumers, learning via social media analytics how they responded to promotions, and increase the relevance of specific stores to shoppers (Chan, 2020:20). Walmart use cloud strategies to build a world-class and worldwide e-commerce platform that enables their customers to purchase goods anywhere and anytime (Chan, 2020:20). Applications of AI within Walmart stores include everything from inventory management to managing consumer interactions (Chan, 2020:20).

# **CHALLENGES AND SOLUTIONS**

Every hour, more than a million shoppers at Walmart purchases goods, which contributes to about 2.5 PB of data collections (Hariri *et al.*, 2019:3). Such enormous data quantities may result in scalability and unpredictability issues, since a database tool may not be able to accommodate infinitely large datasets (Hariri *et al.*, 2019:3). When attempting to scan and comprehend the data at scale using many of the currently available data analysis tools, which are not meant for large-scale databases in any case, they simply fail (Hariri *et al.*, 2019:3). The quick expansion of Walmart required a substantial intake of new workers, and recruiting individuals with the proper talents proved challenging with an analytics operation as big as the one Walmart envisioned (Marr, 2016:9). Walmart turned to the crowdsourcing data science competition website Kaggle as one of its strategies for fixing this issue envisioned (Marr, 2016:9). Users of the website Kaggle were tasked with forecasting the effects of promotional and seasonal events, such as stock-clearance discounts and holidays, on the sales of a variety of goods envisioned (Marr, 2016:9). According to Walmart’s senior recruiter for its Information Systems Operation, Mr. Mandar Thakur: “The Kaggle competition created a buzz about Walmart and our analytics organization. People always knew that Walmart generates and has a lot of data, but the best part was that this let people see how we are using it strategically” (Marr, 2016:10). For the purpose of investigating and implementing fresh data-led projects across the company, Walmart formed @WalmartLabs and their Fast Big Data Team (Marr, 2016:6). The Data Café, a cutting-edge analytics hub at their Bentonville, Arkansas, headquarters, served as the pinnacle of this strategy (Marr, 2016:6). The analytics team at the Café has real-time access to 200 streams of internal and external data, including a 40-petabyte database of all sales transactions over the past weeks (Marr, 2016:6). Real-time data analysis is viewed as essential to boosting company success (Marr, 2016:6). The time it takes from a problem being discovered in the data to a solution being presented has decreased from an average of two to three weeks down to roughly 20 minutes thanks to the Data Café's analytics technologies .

# **DATA TYPES**

The Data Café employs a database that is continuously updated and contains 200 billion rows of transactional data, which only includes the last few weeks' worth of activity (Marr, 2016:8).  Additionally, the Café gathers information from 200 alternate sources, including a database of activities around Walmart stores, meteorological data, economic data, telecoms data, social media data, and data on gas prices (Marr, 2016:8).

# **DATA MIGRATION TOOLS AND SCHEDULES**

The world's largest retailer, Walmart, has been utilizing Google Cloud as part of its innovation journey (Ruth, 2021). The executive vice president, global CTO, and chief development officer, along with his team, developed a strategy to quicken Walmart's transition in three areas: assistance in improving business performance, enhancing customer experiences, and updating their platforms and infrastructure (Ruth, 2021). BigQuery is being applied to some of their most data-intensive and crucial decision-making processes (Ruth, 2021). Google's serverless data warehousing solution is known as BigQuery. Big data from Walmart has moved, and by the end of that fiscal year, the number is expected to nearly double in size (Ruth, 2021). BigQuery offers integration with almost all third-party data virtualization and analytics tools, so users are not restricted to using solely Google solutions (Ruth, 2021). The scalability reduced processing times by 23 percent (Ruth, 2021). That includes having the ability to shut the books on the finances in three days as opposed to five (Ruth, 2021). Google Cloud's AI capabilities has opened the door to new advantages, including the ability to predict demands, manage in-store clouds, optimizing supply chains, and giving employees more time to interact with consumers (Ruth, 2021). Walmart's express delivery service, which went live in 2020, employs AI in the background to optimize delivery routes and decides whether consumers are eligible for the service (Ruth, 2021). As for Walmart’s data migration scheduling, when constructing the Social Genome (also one of Walmart's own software products), Walmart encountered a number of technical challenges, including the volume and speed of data flowing into their Hadoop clusters (Rijmenam, 2013). They built a tool called Muppet since the standard Map-Reduce/Hadoop framework was unable to handle the volume and pace of the incoming data (Rijmenam, 2013). Muppet, which is now open-source, processes data in real-time across all clusters and has the capacity to run many analyses concurrently (Rijmenam, 2013).

# **CONCLUSION**

The competition in business is a factor that can make it thrive (Marr, 2016:10).  Walmart has long been a pioneer in data-driven projects like loyalty and reward programmes, and by fully committing to the most recent developments in real-time, as well as responsive analytics, they have demonstrated their intention to stay a competitive super franchise (Marr, 2016:10).

# **BIBLIOGRAPHY**

Chan, J.O-P. 2020. Digital transformation in the era of big data and cloud computing. *International Journal of Intelligent Information Systems*, 9(3):16-23. International Journal ofIntelligent Information Systems. Date of access: 28 Aug 2022. 10.11648/j.ijiis.20200903.11

Hariri, R.H., Fredericks, E.M. & Bowers, K.M. 2019. Uncertainty in big data analytics: Survey, opportunities, and challenges. *Journal of Big Data*, 6(1):1-16. <https://www.researchgate.net/profile/Reihaneh-Hhariri/publication/333655755_Uncertainty_in_big_data_analytics_survey_opportunities_and_challenges/links/5cfb1b9692851c874c568895/Uncertainty-in-big-data-analytics-survey-opportunities-and-challenges.pdf> Date of access: 28 Aug 2022.

Jha, S., Jha, M., O'Brien, L. & Wells, M. 2014. Integrating legacy system into big data solutions: Time to make the change*.* In. Asia-Pacific World Congress on Computer Science and Engineering. IEEE. pp. 1-10.

Marr, B. 2016. *Big data in practice: How 45 successful companies used big data analytics to deliver extraordinary results*. John Wiley and Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, United Kingdom: John Wiley & Sons.

Rijmenam, M.v. 2013. *Walmart is making big data part of its DNA*. (Big Data). <https://datafloq.com/read/walmart-making-big-data-part-dna/> Date of access: 29 Aug 2022.

Ruth, J-P.S. 2021. *Google cloud next paints digital landscape where data and ai meet*. <https://www.informationweek.com/cloud/google-cloud-next-paints-digital-landscape-where-data-and-ai-meet> Date of access: 29 Aug 2022.