
| RESEARCH ARTICLE

Analysis of Big Data Utilization in Technology Companies (Gojek Case Study: PT GoTo Gojek Tokopedia Tbk)

Usanto S

Institusi Teknologi dan Bisnis SWADHARMA, Indonesia

Corresponding Author: Usanto S, **E-mail:** usanto.s@swadharna.ac.id

| ABSTRACT

Rapid technological developments create trillions of bytes of data every day from various platforms. This amount of data is also called big data, which is often used by various companies in running their business. This study aims to find out how Gojek, under the auspices of PT GoTo Gojek Tokopedia Tbk, utilizes and uses big data. This study uses a literature review research method, while data collection techniques are carried out by exploring journals and other information relevant to the study. The results of the study show that Gojek uses big data to improve business by analyzing personal data such as travel track records, types of food purchased to drug lists. The use of big data is to know individual customers better.

| KEYWORDS

Implementation, Big Data, Technology Company, Gojek

| ARTICLE INFORMATION

ACCEPTED: 20 September 2022

PUBLISHED: 23 September 2022

DOI: 10.32996/jbms.2022.4.4.8

1. Introduction

Information technology continues to grow rapidly, requiring users to take advantage of existing technology. Both users are from young and old age. The existence of the internet has affected several sectors, one of which is online business. In the past, before there was the internet, a company's marketing technique was done door-to-door or by sales marketing that offered one person to another. But now, it is rare for companies to implement marketing strategies as before (Salsabila, nd).

Information technology is an inseparable part of the business world, especially in the face of increasingly competitive business competition. The need for information technology is a basic need for companies to survive in a competitive business world. Information technology has driven advances in product and process technology, as well as the formation of an information society (Indrayani, nd). Through information technology, trillions of bytes of data are created every day from various sources, such as social media, sensors, video surveillance, and smart grids. This sea of data leads to one Big Data terminology (Sirait, 2016). Big data can give companies an advantage with access to a personal understanding of customer preferences, as well as what can please customers. Big data allows this to be done massively and automatically (Rahman, 2021).

Big Data is not a stand-alone technology, technique, or initiative. Big Data is a trend that covers a wide area of business and technology. Big Data refers to technologies and initiatives that involve data that is so diverse, fast-changing, or super-large that it is too difficult for conventional technology, expertise, or infrastructure to capture and deal with it effectively. In other words, Big Data has a size (volume), velocity (velocity), or variety (variety) that is too extreme to be managed by conventional techniques. Big Data involves data creation, storage, information mining, and analysis processes that stand out in terms of volume, velocity, and variety (Pujianto et al., 2018).

However, the use of big data cannot be done directly but requires special tools. This is because the data used is very large, so it cannot be processed using ordinary tools. Big data requires adequate computing with large storage (Hussein, 2022). Based on the

Copyright: © 2022 the Author(s). This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) 4.0 license (<https://creativecommons.org/licenses/by/4.0/>). Published by Al-Kindi Centre for Research and Development, London, United Kingdom.

background of these problems, researchers are interested in conducting research with the title "Analysis of the Use of Big Data in Technology Companies" with a Gojek case study under the auspices of PT GoTo Gojek Tokopedia Tbk. Then, the discussion in this article is related to the theories of implementing big data to answer relevant problems and have their validity tested.

2. Theoretical Basis

2.1 Data

The definition of data is a collection of information or values obtained from observations (observations) of an object. Data is not only numbers, but data can also be symbols or properties. There are many kinds of data, including population data and sample data, observation data, primary data, and secondary data. The usefulness of data is an objective basis that is in a decision-making process in solving problems and making decisions. (Hastono, 2015). Data can be structured data and non-structured data, such as pictures, music, videos, essays, and discussions. When compared to structured data, non-structured data has a better reflection of reality in making a decision (Sutandi, 2018).

2.2 Big Data

Big data is a collection of data whose data is large but has a high diversity of data sources, so the data must be managed in a way or tool with appropriate performance (Maryanto, 2017). Data is obtained from common sources, from applications that he already has on his smartphone, such as social media, this online motorcycle taxi application, and so on. The definition of big data, in general, has five characteristics, namely volume, velocity, variety, value, and the last is veracity. (Toba, 2016). Big Data involves the processes of data creation, data storage, information retrieval, and data analysis that stand out in terms of volume, speed, and variety. (Sedayu & Andriyansah, 2021). Big data, in general, is a type of unstructured data that requires real-time analysis. In addition, big data brings a new opportunity to encourage the creation of new challenges. (Khairani & Irwansyah, 2018).

Big data is a collection of data with large volumes originating from various types of data sources around the world that can be accessed anywhere and anytime and can grow very quickly. The existence of this makes the data owned can be used as an analysis and decision-making tool. One type of big data that can be used is Google Trends. Google Trends is a service provider site to find out how often a topic is discussed, find out what topics are being discussed, and compare performance data between keywords (Mustinda, 2019). Several studies that have been conducted using Google Trends include digital marketing topics (Natasuwarna, 2020), public interest in halal products and halal labels (Nurbaiti, 2019), hotel room occupancy rates (Ayuningtyas & Wirawati, 2021), and tourism (Purnaningrum & Ariqoh, 2019). The number of research topics that have started to use Google Trends shows that Google Trends can be used as an analysis in almost all discussion topics. However, its use in research in the tourism sector in Indonesia is still rarely found, even though the existence of Google Trends can help in tourism development (Purnaningrum & Ariqoh, 2019)

The word Big Data refers to large-scale information management and analysis technologies that exceed the capabilities of traditional data technologies. There are three main differences between Traditional and Big Data, namely the amount of data (volume), the rate of data generation and transmission (velocity), and the types of structured and unstructured data (variety).

2.3 Big Data Analytics Ecosystem

What is Analytics? A starting point for understanding Analytics is a Way to explore/investigate/understand an object down to its roots. Analytics results usually don't cause much confusion, as the context usually makes the meaning clear. Big Data Analytics is an analytical tool and technique that will be very helpful in understanding big data on the condition that the algorithms that are part of these tools must be able to work with large quantities in real-time conditions and on different data.

2.4 Big Data Analytics Tool Ecosystem

Based on the Big Data Ecosystem, which is so complex, the first thing that is very important to do as the key to success in building a Hadoop Ecosystem is to identify the user and how the interactions between, or from each Big Data Tool, and what will be used later in implementing in Hadoop environment. The focus of the ecosystem visualization of the Big Data Tool is related to a system description that studies the interaction between one Big Data Tool with one another, as well as the interaction of an existing Big Data Tool with its environment (e.g., Data, Case Conditions, etc.), while the focus of making Big Data Tool architecture (discussed in more detail in the next chapter) is more of an art and science that is used to design, design and build products from combining the large selection of Big Data Tools available according to how complex the cases are handled for the project analytics in the Big Data Scope, and allows for modifications or improvements to the existing architecture from time to time; for example, in a few months or years, the product architecture may become obsolete and must be replaced with another better and better Hadoop-optimized architecture from Apache has similarities with Linux in terms of composition is, configuration and distribution, but not in terms of functionality. Hadoop composition is built using open-source software under the Apache Foundation license. Hadoop Distribution is a company that makes ready-to-use Hadoop packages and sells them from the results of standard Hadoop configurations with other Big Data tools according to the compositional configuration design that they think is the best, which

includes the Hadoop Ecosystem and Architecture. You can also independently create Hadoop Distribution, but what you have to pay attention to is the basic to advanced skills and mastery of Big Data concepts and the tools that are developing for Big Data; if you haven't, then you can learn them gradually with a focus on several cores of Big Data processing tools, for example, for the prefix it is enough to master Hadoop and Spark, 13 Cholissodin, I., Riyandani, E., 2016, Big Data Analysis, Faculty of Computer Science, Brawijaya University, Malang. Then followed by other tools; this will make it easier for you to learn to become very expert in the field of Big Data and its tools so that after proficiency, you will find it very easy to make Hadoop Distribution independently.

2.5 Big Data Architecture

The best way to get a solution to the Big Data Problem is to "Share the Problem". Big Data solutions can be well understood using Layered Architecture. Layered architecture is divided into different Layers where each layer has specifications for performing certain functions. The architecture helps in designing the Data Pipeline (data path) with various modes, either Batch Processing System or Stream Processing System. This architecture consists of 6 layers that ensure optimal and safe data flow. Data Ingestion Layer - This layer is the first step for data that comes from a particular source and will start its journey. The data here will be prioritized and categorized, so that data can be processed easily and forwarded to the next layer. The tools that can be used are Apache Flume, Apache Nifi (Data Collection and Mining from Twitter using Apache NiFi to Build a Data Lake), and Elastic Logstash. The data enters the Data Lake in raw form, and all the data is stored, not only the data used but also the data that may be used in the future. In Data Lake, all data is stored in its original form.

Data Collector Layer - In this layer, the focus is more on channeling data from the ingestion or ingestion layer to other data paths. In this Layer, the data will be separated according to their groups or components (Topics: user-defined categories whose messages are published, Producers - Producers who post messages in one or more topics, Consumers - subscribe to topics and process posted messages, Brokers - Brokers who are diligent in managing and replicating message data) so that the analytical process can begin. The tool that can be used is Apache Kafka.

Data Processing Layer - the main focus of this layer is on the pipeline data processing system, or we can say that the data that we have collected in the previous layer will be processed in this layer. Here we do the extraction and also learn from the data to be directed to various destinations and classify the data flow that should be directed, and this is the starting point where the analytics has been done. The data pipeline is the main component of Data Integration. The data pipeline flows and transforms real-time data to the services that need it, automates the movement and transformation of data, processes data that runs within your applications, and transforms all incoming data into standard formats that can be used for analysis and visualization. So, a Data pipeline is a series of steps that your data takes. The output of one step in the process becomes the input of the next. The steps of the Data pipeline can include cleaning, transforming, merging, modeling, and more in any combination. The tools that can be used are Apache Sqoop, Apache Storm, Apache Spark, and Apache Flink.

Data Storage Layer (Data Storage Layer) - Storage media becomes a major challenge when the size of the data used becomes very large. This layer focuses on "where to store such large data efficiently". The tools that can be used are Apache Hadoop (HDFS), Gluster file systems (GFS), and Amazon S3.

Data Query Layer - this layer is where the active analytical processing takes place. Here, the main focus is on collecting data values so that they can be made more useful and easy to use for the next layer. The tools that can be used are Apache Hive, Apache (Spark SQL), Amazon Redshift, and Presto.

Data Visualization Layer - The visualization process, or the stage of representing data in a visual form, is probably the most prestigious level, where data pipeline users can experience detailed and easy-to-understand report results from the data values that have been visualized. We need something that will draw people's attention away from the data visualization, thus making your findings easy for them to understand well through the visualization. Find and filter information when you view company data or surf the Internet and don't know where the right information is, "React.js" is a recommender system to predict user criteria, namely determining what kind of user model.

3. Research Methods

This study uses a literature review research method. The literature review research method is a data collection technique by conducting a review study of books, literature, notes, and reports that have to do with the problem being solved. (Danandjaja, 2014). Meanwhile, data collection techniques were carried out by exploring journals and other information relevant to the study. The journals used are searched for in Google Scholar searches with several writing formats, including big data, the use of big data in Gojek, implementation of big data in Gojek, and so on. After being searched and collected, the journals are sorted by year of publication. Journals used as references have standardization, which is a maximum of 10 years after the journal is published.

4. Results and discussion

Gojek is a technology company that has more than twenty products, including transportation and delivery services. In 2016, Gojek became the first Unicorn Company in Indonesia; then, in 2017, it was ranked 17th out of 20 companies that changed the world according to Fortune. Gojek continued to grow rapidly and expand into Vietnam and Thailand, then Gojek merged with the E-commerce company Tokopedia in 2021 (Gojek, 2021).

One of the strategies used by Gojek is through Big Data. Gojek's VP of Data Science, Syafri Bahar, said the Gojek Company collects data related to the security and convenience of partners and customers. All of this data is then analyzed further to develop new features or services that are useful for users (Annur, 2020). The data collected is in the form of personal data that has been described and approved by the user on the Gojek Privacy Policy page. The personal data contains a travel track record of the type of food purchased, the grocery list, the type of medicine used, the massage schedule, and the vehicle washing schedule. Creative economy fields such as music and film can take advantage of big data technology. One example of the use of big data in business is to make companies know their customers better (Cerdiknya. co, 2021).



Figure 1. The Substance of Big Data

The big data ecosystem in Gojek can be analogized as described by Nadiem; for example, there is an active Go-Food user in Kemang who often orders KFC fried chicken, Crepes, and Bu Nani's fried bananas. Other customers who live in Kuningan also frequently order these three types of food. Then one time, customers in Kemang ordered new types of food and ordered quite often. The activity of one type of food that is quite frequent indicates someone likes that food. Since the customers in Kemang like the new type of food, it is likely that the customers in Kuningan also like it because they have shown a similar pattern of tastes. In addition, Nadiem also added that if an application can suggest a new food with a percentage of 90% of customers who will like it, customers will consider them as friends and be reluctant to switch to other applications (Gabriel, nd). This is one of the efforts targeting consumer loyalty.

Consumer loyalty is a deeply held commitment to repurchase or patronize a preferred service product consistently in the future, regardless of situational influences and marketing efforts making the potential to cause switching behavior (Ramdani, 2020). According to Griffin, consumer loyalty is more associated with behavior (behavior) than with attitude. If someone is a loyal consumer, the consumer will show buying behavior which is defined as a non-random purchase that is expressed from time to time by several decision-making units (IRWANDI, 2019). Two important conditions related to loyalty are 1. Customer retention (customer retention) describes the length of the relationship with the customer. Customer retention rate is the percentage of customers who have fulfilled several repeat purchases over a limited period. 2. The total share of customers (total share of customers) of a company shows the percentage of the customer's budget spent on the company (Dwi Wahyuni, 2017).

5. Conclusion

Big data can give companies an advantage with access to a personal understanding of customer preferences, as well as what can please customers. However, the use of big data cannot be done directly but requires special tools. This is because the data used is very large, so it cannot be processed using ordinary tools. Big data requires sufficient computing with large storage. One of the strategies used by Gojek is through Big Data. Gojek collects data related to the security and convenience of partners and customers. All of the data is then analyzed further to develop new features or services that are useful for users. The personal data contains a travel track record of the type of food purchased, the grocery list, the type of medicine used, the massage schedule, and the vehicle washing schedule.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers.

References

- [1] Annur, CM (2020). *How Gojek, Telkom & Deloitte Take Advantage of Big Data amid a Pandemic*. Data said.
- [2] Danandjaja. (2014). *Library Research Methods. Indonesian Anthropology*.
- [3] Dwi-Wahyuni, CR (2017). *Service Quality And Its Influence On Customer Loyalty Mediated By Satisfaction At Bank Muamalat Jombang*. *Exist Journal of Economics and Business Research*, 12 (1), 69–82. <https://doi.org/10.26533/eksis.v12i1.84>
- [4] Gabriel, J. (nd). *Peeking into the Big Data Ecosystem at Gojek*. *Sriwijaya University*.
- [5] Hussein, S. (2022). *Understanding Big Data: Technology, Analysis, Strengths, and Weaknesses*.
- [6] Indrayani, H. (nd). *APPLICATION OF INFORMATION TECHNOLOGY IN INCREASING COMPANY EFFECTIVENESS, EFFICIENCY, AND PRODUCTIVITY*.
- [7] IRWANDI, A. (2019). *The Influence of Minimarket Image and Service Quality on Satisfaction and Its Implications on Indomart Customer Loyalty in the Cirebon City area*. Unpas Library.
- [8] *Milestones*. (2021). Gojek.
- [9] *Utilization of Big Data in the Gojek Business*. (2021). Clever. Co.
- [10] Pujianto, A., Mulyati, A., & Novaria, R. (2018). *UTILIZATION OF BIG DATA AND PROTECTION OF PRIVACY*. *BIJAK Scientific Magazine*, 15 (2), 127–137.
- [11] Rahman, BA (2021). *These are the benefits of big data for companies*. Indonesian Cloud Computing.
- [12] Ramdani, R. (2020). *The Effect of Product Quality And Trust In Consumer Loyalty In Elzatta Products, Bandung Indah Plaza Branch*. Indonesian Computer University.
- [13] Salsabila, R. (nd). *Benefits of Using Big Data for Online Ojek Business*. *Benefits of Using Big Data for Online Ojek Business*.
- [14] Sirait, ERE (2016). *Journal of Postal and Informatics Research IMPLEMENTATION OF BIG DATA TECHNOLOGY IN INDONESIAN GOVERNMENT INSTITUTIONS IMPLEMENTATION OF BIG DATA TECHNOLOGY IN GOVERNMENT INSTITUTIONS IN INDONESIA*. *Journal of Postal and Informatics Research*, 6 (2), 113–136. <https://doi.org/10.17933/jppi.2016.060201>