The Use of Big Data in Supporting Customer Profiling

Gunawan Wang[1], Lazuardi Ridho Maulana[2], Nico Leonardi[3], Emil Robert Kaburuan[4]
[1,2,3]Information System Management Department, BINUS Graduate Program – Master of Information System
Bina Nusantara University, Jakarta, Indonesia 11480
[1]gwang@binus.edu; [2]lazuardi.maulana@binus.ac.id, [3]nico.leonardi@binus.ac.id, [4]emil.kaburuan@binus.edu

ABSTRACT

Nowadays, the online loan business is growing rapidly. This kind of business offers a low limited loan amount for customer needs. The online loan business only verifies the current data of each customer for the user profiling process. Due to insufficient profile data of customers in person, the limit cannot go higher than it is now. But this business has the potential to grow and provide a better-customized loan amount according to each customer's financial conditions. The lack of customer data profile can be solved by integrating the data and loan system with big data. Big data offers a way to give online loan business a more accurate profiling system to track the customer financial history and many more. By having sufficient data of potential customers, the online loan companies can provide higher loan amounts to the customers based on their financial track.

Key words : big data, business, loan, online, profiling system

1. INTRODUCTION

Online loan business growth shows that this type of business has a lot of potential in Indonesia. There are many startup companies that are using this concept of the online loan provider to gain profit. However, the maximum amount of loan provided is very low. It is around 20 million Rupiah at maximum. It is far less than the average amount of down payment needed to buy a property, such as a house for a homestay that is around 40 to 50 million Rupiah, if, for example, a customer wants to buy a house in South Tangerang area. Beside of that, customer profiling methods used in this business is still not optimal and the data obtained from customer are very limited.

There is no option to get a loan as high as 50 million rupiahs, no matter if a customer can afford it based on his or her financial track. What are the things that are needed by the loan provider companies to be brave enough to give a customer a huge loan as high as 50 million rupiah or more? It is strong bases that the customer can afford to return the loan and the loan interest in a specified period.

How do the companies get strong proof that a customer really able to return the loan and the interest in a specified period? It is by analyzing the financial historical data of a customer. This data can be obtained if a system can integrate with a customer personal data, such as social media, fintech accounts, e-commerce history, etc.

This great amount of data can be processed with big data. By using big data, a system can process a lot of data to cluster all data and do the customer profiling. Big data has the ability to handle a massive number of data spread across social media, mass media, and other media.

2. LITERATURE REVIEW

2.1 Financial Services

Financial services are a business that services that helping the process of financial goods until it acquired but it also covers a broad spectrum of investment funding, including securities[1]. Financial services can be brought with many financial industries like banks and insurance companies, but nowadays the financial services are popular with fintech who serve the exact same financial services but the more faster process of validation and aiming to get better customer experience[2]. With this newborn of the phenomenon, financial services from fintech struggling with the regulation of technological change that slowly in certain countries circumstance[1].

2.2 Fintech

Fintech comes from the term financial technology. According to the National Digital Research Center (NDRC), in Dublin, Ireland, defining cutting-edge technology as "financial services of innovation" or "innovation in financial services fintech" 2, which is an innovation in the financial sector that obtains the Touch of modern technology. Financial transactions through this technology cover payments, investments, money loans, transfers, financial plans and comparison of financial products. There are currently 142 companies involved in fintech identified[3]
The financial technology industry (fintech) is one of the financial services methods that is gaining popularity in the

current digital era. And digital payments are one of the sectors in the most developed FinTech industry in Indonesia. The sector and the community expect this sector to encourage an increase in the number of people who have access to financial services.

2.3 Big Data

The growth of communication and information technology today is encouraging to be able to process structured and unstructured data [4]. Structured or unstructured data processing can be processed using a variety of platforms, one of the best technology is big data. In general, big data is data with high volume, rapid growth and varied information assets that are expected to help the processing of affective and innovative information to improve perception and decision making [5]. Big data is an important tool for professionals industry to process structural, semi-structural and unstructured data.

Applications based on big data in property business can be divided into two parts, namely front-end and back-end. The front-end applications of big data are business intelligence that consists of several panels and functions to display data according to the trends of property and people's intention. The business intelligence in the property business has a variety of data sources that are not completely supplied by big data technology, but there are still conventional database management systems (DBMS), namely Oracle, Microsoft SQL Server, MySQL, and Sybase. The back-end in big data applications are engines can process unstructured data such as social media images, video, or internal application service applications or business partners. Structured data provided by conventional DBMS will be processed and combined with unstructured data. Processing results will be used in reports or stored in a data warehouse for later use in strategic analytical needs.

Big Data is data that has the scale, distribution, diversity, and use of the latest architecture with a combination of several new techniques to allow information that opens up business value. Big Data has three main characteristics: Huge volume of data, Complexity of data types and structures, and Speed of new data creation and growth [6]. An example given by Thomas Freeman in his book World is Flat (2005), Big Data technology has been adopted in several industries to store a collection of EDI (Electronic Data Interchange) records, integrated information systems, as well as various types of data and metadata that can be proven or credible. The technology that supports data integration through the most used middleware is SOA. With the integration of data, it will further enrich the data sources of the big data for the real estate business.

The current property business needs fast and accurate information to support every decision that will be taken to maximize the competencies of the main business of the organization. These competencies are things that are very

good for organizations and that can guarantee their survival in the future, provided that organizations can develop them to continue complying with market requirements [7]. The problem for each organization is to analyze the situation in the market with valid data to support each step that is taken, business intelligence works as a support tool in the analysis of data for use in decision making. According to Laursen & Thorlund, Business Intelligence is a tool to analyze business and provide information in the form of reports for the right people at the right time[7]. Business intelligence is one of the current tools that can help provide information to companies with their ability to process data to be able to see the operational process of organizations or trends that are currently in society, the ability to display information in various ways, such as charts, monitoring dashboards, daily reports. Business intelligence mainly use four main techniques to analyzing the data that can be used in data mining which are classification, prediction, association and clustering[11]. The power of business intelligence that various organizations are looking for. With the development of IoT and big data, the data source for business intelligence is expanding because the capabilities of each of these technologies greatly help business intelligence to process information more validly and quickly.

Applications based on big data in property business can be divided into two parts, namely front-end and back-end. The front-end applications of big data are business intelligence that consists of several panels and functions to display data according to the trends of property and people's intention. The business intelligence in the property business has a variety of data sources that are not completely supplied by big data technology, but there are still conventional database management systems (DBMS), namely Oracle, Microsoft SQL Server, MySQL, and Sybase. The back-end in big data applications are engines can process unstructured data such as social media images, video, or internal application service applications or business partners. Structured data provided by conventional DBMS will be processed and combined with unstructured data. Processing results will be used in reports or stored in a data warehouse for later use in strategic analytical needs.

3. METHODOLOGY

The author uses a case study method to analyze the current infrastructure of big data that can help startup fintech who have loan services better at profiling and scoring the risk. In Indonesia where the startup has to compete against the conventional bank to serving good financial services and one of the most anticipated on the market loan. We gather some information on their official website and ads to compiling the general information about the services, in this gathering information phase we collect the information on how to applying this kind of service. After we collect some general

information of eligible fintech to operate in Indonesia financial authority (OJK) website. We can assume that most of fintech need NIK, Mobile Phone Number, and email. However, the fintech is necessary to accept the media and cookies of the web browser, so It leads to more unstructured data like XML, response log, image or photos, and so on. Besides cloud base infrastructure services (storage, compute, infrastructure/VM management) the following specific applications and services will be required to support Big Data and other data-centric applications[8]. In paper who describe by Deschemko in general big data infrastructure mainly has Big Data Management tools, Registries, indexing/search, semantics, namespaces, Security infrastructure (access control, policy enforcement, confidentiality, trust, availability, privacy), Collaborative environment (groups management).

This must-have element of general big data can be a core of how to managing how the data is collected until it visualizes into something that people can make a decision with. It will need more resources in infrastructure to analyze after gathering all the information into big data, then the analytic to make information more valuable is necessary. We manage to know that big data can possibly improve with such a decent infrastructure for Cluster services, Hadoop related services, and tools, Specialist data analytics tools (logs, events, data mining, etc.), Databases/Servers SQL, NoSQL, MPP (Massively Parallel Processing) databases [9], managing all this features using the cloud PaaS and IaaS is clearly one of the option to begin with, another option using third party support who can manage cloud infrastructure with specialization in taking care of everything such as data hosting, data sharing, file backup-storage as well as productivity software handling service[10].in this paper we will discuss why does the fintech need to build everything or just part of the infrastructure.

4. DISCUSSION

4.1. How it works

A. Integration with Data Sources

When a customer applies the loan request, they are required to give access permission to their emails, phone numbers, online accounts such as social media, e-commerce accounts, bank accounts (account balance, tax history, etc), and enters the property info they are targeting on. Besides personal data, the company system is also needed to get information from other sources such as web, news, etc for better info about property and area by using API management system to integrate the data from all sources to the system.

B. Big Data Processing

After the system has been integrated with all the sources needed. Big data will have to process the data to analyze or do the profiling of a customer's financial data with Map Reduce (MR). The MR programming structure can improve the complicated nature by running parallel information [12]. By checking all data in the social media and e-commerce accounts, a system may know a person's lifestyle, expenses, incomes, and many more. By checking all bank accounts balance, tax transaction history, personal assets, a system may know how stable a person's financial status, by analyzing the web or new, a system may predict if the target property's value will increase by time or decrease, which can be used for property investment.

For clustering the big data, Hadoop can be used as an option to cluster these massive data for profiling each customer accurately. In this current real-world situation, big data can extracting, transforming, and loading any type of data into big data, with the current infrastructure and supporting technology that can be drawn general infrastructure of big data in Figure 1.
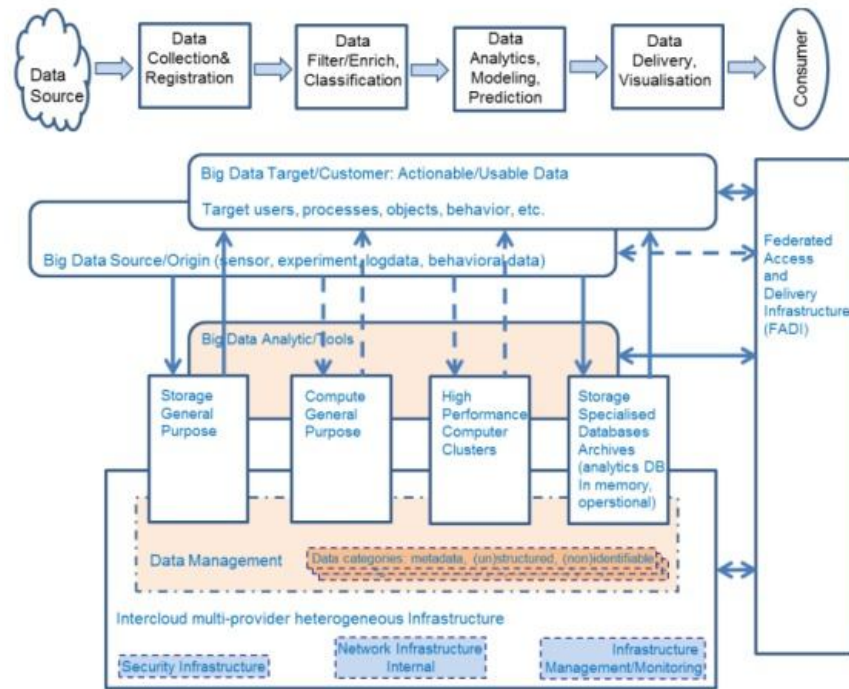
Figure 1: General Big Data Architecture & Flow

As shown in Figure 1, this is how data will be processed, this is the general infrastructure for general data management, typically cloud-based, and Big Data Analytics part that will require high-performance computing clusters, which in their own turn will require high-performance low-latency network. first, all data sources from social media, e-commerce account, bank account, etc will be collected and will be registered to the system. Then the data will be filtered, enriched and classified into certain categories. After that, the data can be modeled, used for predicting, and general analyzing. Thus, a big data analytics tool can be used to get information from the data. Apache Hadoop and Apache Cassandra can be the option to be used for big data implementation.

Apache Hadoop has HDFS (Hadoop Distributed File System) which is the ability to hold all type of data such as videos, images, JSON, XML, and plain text over the same file system, highly useful for R&D purposes, provides quick access to data, highly scalable, and highly-available service resting on a cluster of computers.

Apache Cassandra is a distributed type database to manage a large set of data across the servers which have capabilities such as continuous availability as a data source, linear scalable performance, simple operations, across the data centers easy distribution of data, cloud availability points, scalability, and good performance.

The use of big data may change the course of fintech to a new way of analyzing the data, the use of cloud-based big data that can simplify the building of such infrastructure and provision it on-demand. The use of cloud computing technology in big data can make the efficient operation of such infrastructure will require both overall infrastructure management and individual services and infrastructure segments to interact between themselves[9]. Analyzing the data from a lot of sources and processing into valuable information in cloud infrastructure in terms of knowing what are customer behave is the power to profiling each one of their respective customers. With cloud computing, it will easier to maintain and virtualize the infrastructure provided, with help of cloud computing in big data environment it might extract every unstructured data from social media, email, eCommerce, cookies, and so on, with integration of API Management to bridging another apps into big data environment and continue the process with ETL Technology and stored into side or outside organization into Datawarehouse for processed into a chunk of information. Cloud environment needs to make sure they have capable network infrastructure and security infrastructure whose managing the traffic of data and their encryption, this infrastructure includes the load balancing, encryption machine, file gateway, and firewall. In Figure 2 we can see the infrastructure is capable of integrating across network providers to hit every virtual cloud server.
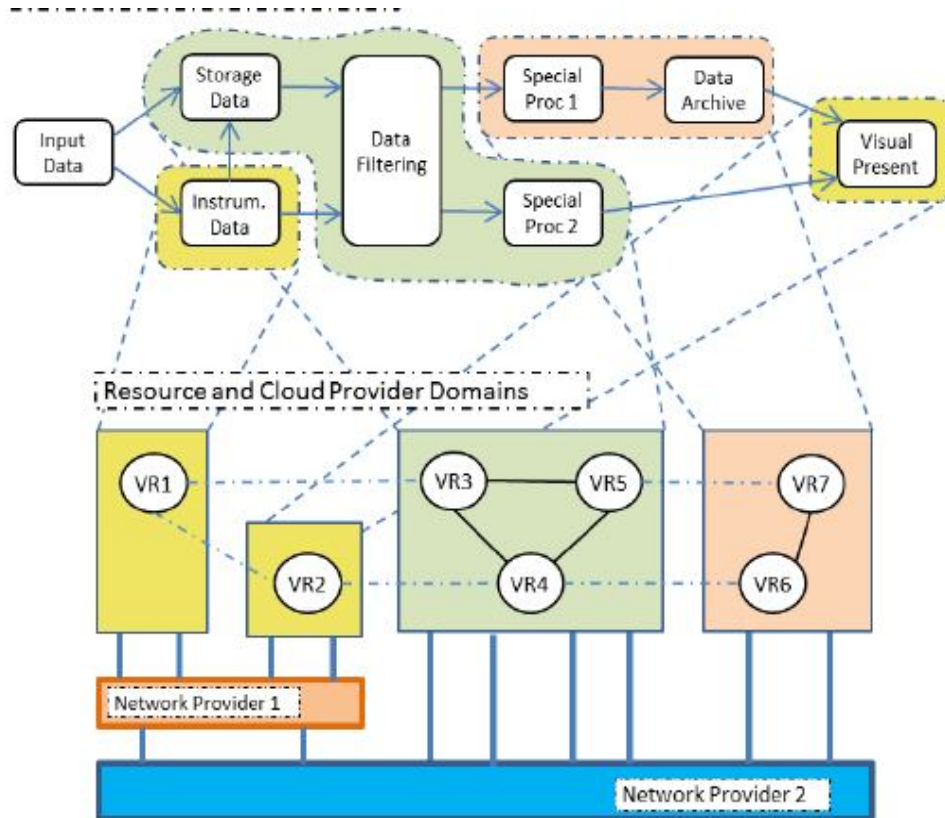
Figure 2: Big Data Cloud Computing-based

C. Gaining Analyzing Process Results

After big data clustered all data streams and produced detailed information of a customer from a big data analytics tool, the system can show a prediction of a customer with credit scoring by the result of data analytics. As a result of this paper finds, combined technology using the general function of big data and cloud computing is come into best of breed, Using the public cloud environment to interconnecting user and system with API then transforming all the data with SH1 or MD5 encryption inside the hardware security module and all the traffic is manage by f5 as load balancing and traffic management. Besides, the Data analytic can be placed inside the virtual server or physical server that have proper technology of in-memory and Distributed File System to make the collection of data, extracting the data, transform the data, and visualize can make the process faster. For the credit scoring our big data environment should be integrated by API Management to communicate with Bank Indonesia System and business partner that open the data source of historical financial behavior.

4.2. Opportunities and Advantages

The use of big data in business is still low in Indonesia. By using big data, a company can get more data to be considered for lending a loan to a customer. E-loan company in the present time only has few of customer data such as identity card, a copy of bank account balance record for the last 3 to 6 months, and salary check, while by using a system that is integrated to many customer online accounts and big data tools, a company can have more data to be considered and will be able to give a better-personalized lend recommendation to each customer based on customer personal finance condition.

The property business in Indonesia, especially in Jakarta, Banten, and West Java Provinces is growing. There are a lot of property companies expand their area to the outer side of a big city such as Pasar Kemis, Cikupa, Bintaro, Cilejit and many more. This expansion will increase the supply of property that can be a target for the users. Besides that, this loaning concept is still fresh in Indonesia and the number of competitors is still low. The loaning company will be able to have a better knowledge of each customer and the business potential is high.

4.3. Future Research

The authors suggest future research about the compatibility of this loaning concept to Indonesian regulation or law and

data security and also research to design how the data will be processed in detail to get accurate analyzation result by using business intelligence.

5. CONCLUSION

Big data is necessary for the current industry, especially in financial services. Because every person has some unique financial behavior, this condition force fintech to capture every data in many forms into valuable information. Fintech needs to adapt a proper big data architecture for their organization to enhance its data transformation to be more effective. But fintech needs to be more careful of how the feature used in their big data environment, they should concern about the security and source management. Furthermore, a big data environment can be integrated across organizations to get more data and produce more accurate in the analytic result.

REFERENCES

1. I. Asmundson, "Financial Services: Getting the Goods," 2018. [Online]. Available: https://www.imf.org/external/pubs/ft/fandd/basics/finserv.htm.

2. P. Gomber, R. J. Kauffman, C. Parker, and B. W. Weber, "On the fintech revolution: interpreting the forces of innovation, disruption, and transformation in financial services," *J. Manag. Inf. Syst.*, vol. 35, no. 1, pp. 220–265, 2018. https://doi.org/10.1080/07421222.2018.1440766

3. I. Muzdalifa, I. A. Rahma, and B. G. Novalia, "Peran Fintech Dalam Meningkatkan Keuangan Inklusif Pada UMKM Di Indonesia (Pendekatan Keuangan Syariah)," *J. Masharif al-Syariah J. Ekon. dan Perbank. Syariah*, vol. 3, no. 1, 2018.

4. N. Kshetri, "Big data′s impact on privacy, security and consumer welfare," *Telecomm. Policy*, vol. 38, no. 11, pp. 1134–1145, 2014. https://doi.org/10.1016/j.telpol.2014.10.002

5. Gartner, "Big data," 2019. [Online]. Available: http://www.gartner.com/it-glossary/big-data/·.

6. David Dietrich, Barry Heller, and Beibei Yang, *Data Science and Big Data Analytic*, First Edit. Indianapolis: John Wiley and Sons.Inc, 2015.

7. G. H. N. Laursen and J. Thorlund, *Business analytics for managers: Taking business intelligence beyond reporting*. John Wiley & Sons, 2016. https://doi.org/10.1002/9781119302490

8. Y. Demchenko, C. De Laat, and P. Membrey, "Defining Architecture Components of the Big Data Ecosystem," pp. 104–112, 2014.

9. M. Turk, "A chart of the big data ecosystem, take 2." October, 2012.

10. F. H. Ayob, P. Jayatis, and S. Subramaniam, "The Optimization of Resources Allocation to Minimize Energy Consumption Within Data Center in Cloud," IJATSCE, vol. 8, no. 1, 2019. https://doi.org/10.30534/ijatcse/2019/5281.42019

11. N. Isa, N. Syuhada, M. Yusof, M. A. Ramlan, and U. T. Mara, "The Implementation of Data Mining Techniques for Sales Analysis using Daily Sales Data," IJATSCE,vol. 8, no. 1, pp. 1–7, 2019. https://doi.org/10.30534/ijatcse/2019/1681.52019

12. Ayob, Fahrul Hakim Jayatis, Punaya Subramaniam, Shamala, ″The Optimization of Resources Allocation to Minimize Energy Consumption Within Data Center in Cloud,″ IJATCSE, Volume 8, No.5, September - October 2019. https://doi.org/10.30534/ijatcse/2019/5281.42019