

**Ideation Phase**  
**Literature Survey**

Date	19 September 2022
Team ID	PNT2022TMID15445
Project Title	Global Sales Data Analytics
Maximum Marks	2 Marks

**[1] PERFORMANCE ANALYSIS OF SALES BIG DATA PROCESSING USING HADOOP AND HIVE  
IN CLOUD ENVIRONMENT - 2020**

**(Hanindia Prami Swari, Kadek Susila Satwika, Putu Susila Handika)**

Big data is receiving a lot of attention these days from academia and the IT sector. This is a result of the incredible rate at which data is currently growing, which necessitates the use of numerous qualified data storage and processing solutions in order to meet the 5 V's challenge of big data. The purpose of this study is to evaluate how well large data processing performs. On the cloud server, the sales data will be processed in parallel and controlled with Hadoop and Hive. According to the study, the number of virtual machines deployed has an inverse relationship with the amount of processing time needed. However, according to the results of the block size tests, increasing the block size from 2 MB to 4 MB or 8 MB greatly reduces the time required to execute a query, however increasing the block size from 4 MB to 8 MB has no impact on this time. Due to factors such as the fast growing population data growth is accelerating day by day. The kinds, formats, semantics, organizations, granularity, and accessibility of many datasets vary to varying degrees. Big data has many components, including data representation, which tries to make data more valuable for computer analysis and user understanding. Companies now produce more data than ever thanks to IoT, cloud computing, and artificial intelligence. Social media and the internet both play a huge role in making circumstances unsettling. Accordingly, there are more than 350.000 tweets sent every minute, 500 million tweets sent each day, and over 200 billion tweets sent year. The use of apps and web servers generates an excessive volume of data logs in practically all enterprises.

**[2] WALMART'S SALES DATA ANALYSIS- A BIG DATA ANALYTICS PERSPECTIVE – 2019**  
**(Manpreet Singh, Bhawick Ghutla, Reuben Lilo Jnr, Aesaan F S Mohammed, Mahmood A Rashid)**

In the twenty-first century, information technology is taking off, generating vast amounts of data that must be processed and researched in order to make sense of it as the conventional method is no longer efficient. Retailers now require a 360-degree perspective of their customers in order to avoid missing the market's competitive edge and to achieve their sales and marketing objectives, retailers must develop strong promos and offers. Because their retail sites are spread out geographically, retailers frequently find it difficult to understand the state of the market. Furthermore, it provides merchants with useful and analytical

information, particularly when identifying customers who are in a specific store at a specific time and located in various geographic regions. We analyzed the data sets of Walmart Store, one of the largest retailers in the world, to determine how these scenarios will affect sales at various stores. Walmart's marketing initiatives, and their data-driven business strategies through the visual depiction of the processed data. Businesses need to reconsider the new ways to a better understanding of the customers in this technology era of large scale data in order to acquire a competitive edge in the market. Business analytics would assist build a picture by using data visualization to give retailers insights into their business. With the help of these insights, firms can adjust their strategy in the future to optimize earnings and success. The majority of raw data, especially large-scale datasets, is not useful in its unprocessed form. We can extract significant insights from this collection of bits by employing the proper set of tools. The primary goal of this study is to examine and analyze the datasets that Walmart has made available in order to provide insights and a broad picture of the corporation. These insights can assist merchants in better understanding market circumstances and the numerous aspects influencing sales, such as how the Easter vacation would cause a jump in sales. The above criteria are used to determine how to observe customer wants.

### **[3] Analytics-driven solutions for customer targeting and salesforce allocation - 2007 (R. Lawrence C. Perl ich)**

Sales professionals must find new sales prospects, and sales leaders must assign the sales force to the accounts with the highest likelihood of generating future sales. We outline two analytics-based solutions created by IBM to deal with these relevant problems. With the help of the web-based tool On TARGET, you can find fresh sales prospects at both existing client accounts and non-client businesses. At the level of the product and brand, the algorithms calculate the likelihood of purchase. They extract explanatory features from transactional data along with corporate firmographic data and train on previous transaction samples (e.g., revenue and number of employees). Sales representatives struggle to become experts in particular products or markets (unless the organization commits to a specialized sales force allocated by geography).

### **[4] DATA-DRIVEN SALES LEADS PREDICTION FOR EVERYTHING AS A SERVICE IN THE CLOUD - 2022 (Chul Sung, Bo Zhang, Chunhui Y. Higgins Yoonsuck Choe)**

A cloud platform website that offers a selection of services works with a freemium or free trial business model and actively markets to past visitors. However, due to the following difficulties, there are a number of limitations with current methods. In this study, we present a cloud everything as a service (XaaS) paradigm for data-driven iterative sales lead prediction, which includes cloud platforms and applications. In this framework, we work together to gather business intelligence from various business stakeholders using the BizDevOps methodology. Our RFDL (Recency, Frequency, Duration, and Lifetime) analysis is used to build service usage scores from these business insights, and we estimate sales lead forecast based on the usage scores in a supervised manner. Through multiple iterations of the entire process,

our framework retains its performance of sales lead prediction and, in the end, effectively communicates the prediction results to the sales or marketing team. We anticipate that our iterative, scalable method to sales lead prediction will find broad application in online or cloud business areas where consumer traffic is constantly fluctuating. Businesses can track and analyze client activity more rapidly and intelligently thanks to cloud services. Instead of simply collecting data, we must understand what data we will collect and how customers will use our cloud services if we are to give the correct data-driven business insights. Collaboration between teams made up of different stakeholders is essential if consumer behavior is to be understood adequately. User behavior analysis is carried out by the business's digital marketing team, a data-driven sales lead analysis is done by the data science team, and a sales lead analysis is done by the sales team.

**[5] Analysis and Optimization of Online Sales of Products - 2017**  
**(Z. Pirani, A. Marewar, Z. Bhavnagarwala and M. Kamble)**

Given the massive amount of e-commerce websites currently available, it is essential to have analysis tools that can help assess whether a firm is accomplishing its intended sales goals. Our research discusses the need for a system to evaluate the database transactions of e-commerce websites using various data mining techniques and algorithms, such as affinity analysis, logistic regression, and linear regression, with the goal of producing a "Sales Analytics Tool." Through the development of a system that inputs database transactions of sold products, splits the data collected, examines the graphs, and extracts market trends and product sales patterns is our proposed notion in this research study. The system then refines this data based on market demands, enhancing sales and merchandise planning and ultimately raising the organization's total productivity and revenues.