

ROLE OF BIG DATA IN SUPPLY CHAIN MANAGEMENT
A DISERTATION
SUBMITTED IN THE PARTIAL FULFILMENT OF THE REQUIREMENTS
FOR THE AWARD OF DEGREE
OF
MASTER OF TECHNOLOGY
IN
PRODUCTION AND INDUSTRIAL ENGG.

Submitted by:-

Vikrant Sharma

Roll no.: 2K14/PIE/19

Under the supervision of

DR. N. YUVRAJ



**DEPARTMENT OF MECHANICAL, PRODUCTION & INDUSTRIAL
ENGINEERING**

DELHI TECHNOLOGICAL UNIVERSITY

(formerly Delhi College of Engineering)

Bawana Road, Delhi- 110042

MAY,19

DELHI TECHNOLOGICAL UNIVERSITY

(formerly Delhi College of Engineering)

Bawana Road, Delhi- 110042

CANDIDATE'S DECLARATION

I, Vikrant Sharma, Roll No. 2K14/PIE/19 student of M.Tech (Production and Industrial Engineering), hereby declare that the project Dissertation titled "Role of Big Data Analytics in Supply Chain Management" which is submitted by me to the Department of Mechanical Proction and Automobile Engineering, Delhi Technological University, Delhi in partial fulfilment of the requirement for the award of degree Master of Technology is original and not copied from any source without proper citation. This work has not previously formed the basis for the award of any Degree, Diploma Associateship , Fellowship or other similar title recognition.

Place: Delhi

(VIKRANT SHARMA)

Date: MAY '19

**DEPARTMENT OF MECHANICAL, PRODUCTION & AUTOMOBILE
ENGINEERING**

DELHI TECHNOLOGICAL UNIVERSITY

(formerly Delhi college of Engineering)

Bawana Road ,Delhi-110042

CERTIFICATE

I hereby certify that the Project Dissertation titled “Role Of Big Data Analytics In Supply Chain Management” which is submitted by Vikrant Sharma, 2K14/PIE/19 (Department Of Mechanical Engineering) , Delhi Technological University, Delhi in partial fulfillment of the requirement for the award of the Degree of Master of Technology, is a record of the project work carried out by the student under my supervision. To the best of my knowledge this work has not been submitted in part or full for any Degree or diploma to this University or elsewhere.

Place: Delhi

(DR. N. YUVRAJ)

SUPERVISOR

DATE: MAY '19

Contents

Cover.....	
Candidate's Declaration.....	
Contents.....	
Abstract.....	
Introduction.....	
Literature Review.....	
CH.1 BUSINESS DRIVERS FOR BIG DATA.....	
1.2 UNDERSTANDING THE BUSINESS DRIVERS	
1.3 LOWERING THE BARRIER TO ENTRY	
1.4 CONSIDERATION	
CH. 2 BUSINESS PROBLEMS SUITED TO BIG DATA.....	
2.1 VALIDATING (AGAINST) THE HYPE: ORGANIZATIONAL FITNESS	
2.2 THE PROMOTION OF THE VALUE OF BIG DATA	
2.3 BIG DATA USE CASES	
2.4 TYPICAL QUALITIES OF BIG DATA USAGE	
2.5 VALUE OF BIG DATA USE	
2.6 FORWARD THINKING ABOUT VALUE	
CH. PREPARING ORGANIZATION FOR BIG DATA.....	
IMPORTANT QUESTIONS	
3.2 THE ANCIENT ANGLE TO REPORTING AND PROCESSING	
3.3 THE WAY OF LIFE CLASH ASSIGNMENT	

3.4 THINKING ABOUT ASPECTS OF ADOPTING BIG DATA

CH.4 TAKING CHOICES ABOUT USE OF BIG DATA TECHNOLOGIES.....

4.2 THE STRATEGY FOR BIG DATA IMPLEMENTATION

4.3 STANDARDIZE PRACTICES FOR SOLICITING BUSINESS CUSTOMER EXPECTATIONS

4.4 ACCEPTABILITY STANDARDS FOR BIG DATA STRUCTURE

4.5 MAKING READY THE ENVIRONMENT FOR MASSIVE SCALABILITY

4.6 PROMOTE STATISTIC REUSE

4.7 INSTITUTE PROPER STAGES OF OVERSIGHT AND GOVERNANCE

4.8 PROVIDE A GOVERNED SYSTEM FOR MAINSTREAMING TECHNOLOGY

4.9 CONCERNS FOR COMPANY INTEGRATION

CH.5 RECORDS GOVERNANCE FOR BIG DATA PROCESSING.....

5.1 THE EVOLUTION OF DATA GOVERNANCE

5.2 BIG DATA AND DATA GOVERNANCE

5.3 THE DIFFERENCE WITH BIG DATASETS

5.4 MASSIVE DATA OVERSIGHT: FIVE KEY CONCEPTS

5.4.1 DEALING WITH CUSTOMER RECORDS EXPECTANCIES

5.4.2 IDENTIFYING THE DIMENSIONS OF RECORDS

5.4.3 CONSISTENCY OF METADATA AND REFERENCE INFORMATION FOR ENTITY EXTRACTION

5.4.4 REPURPOSING AND REINTERPRETATION

5.4.5 INFORMATION ENRICHMENT AND ENHANCEMENT

5.5 CONCOR

CH 6.1 USE CASES.....

6.2 CONCERNS: INFRASTRUCTURE BEDROCK FOR THE INFORMATION LIFECYCLE

6.3 BIG DATA APPLIANCES: HARDWARE AND SOFTWARE PROGRAM TUNED FOR PROCESSING

6.4 ARCHITECTURAL SELECTIONS

6.5 THINKING ABOUT PERFORMANCE TRAITS

6.6 ROW VERSUS COLUMN-ORIENTED FACTS LAYOUTS AND ALERTNESS PERFORMANCE

6.7 CONSIDERING PLATFORM ALTERNATIVES

CH.7 BIG DATA EQUIPMENT AND TECHNIQUES.....

7.1 ABOUT BIG DATA STORAGE

7.2 A WIDESPREAD REVIEW OF LARGE PERFORMANCE ARCHITECTURE

7.3 HDFS

7.4 MAPREDUCE AND YARN

7.5 INCREASING THE BIG DATA SOFTWARE ENVIRONMENT

7.6 ZOOKEEPER

7.7 HBASE

7.8 HIVE

7.9 PIG

CH.8 CONCLUSION AND DISCUSSION

References

ABSTRACT

The aim of this research work is to highlight the role Big Data Analytics can play in improving the performance of Supply Chains and businesses. Any business which wants to maximize its profits by making optimum use of the resources available can use this research work for gaining insight into implementation of Big Data Analytics for the same. Research work done by scholars worldwide has been thoroughly analysed and personal research work has also been done by the author to prepare this report. It is hoped that this work will help researchers in the future for carrying forward the research work in the Big Data Analytics field for the benefit of businesses. Big Data can improve supply chain management a lot. It can be used in medical field, government policy- making functions and many other fields. It has applications in both goods and services sector. Processing of Big Data can help in decision making and increases profits. Large studies done by researchers have revealed that companies which employ Big Data analysis techniques are more efficient than their competitors. Data Analysts are very much in demand now-a-days. To be a good data analyst, one must have industry-specific knowledge also apart from being good in data analysis techniques. What separates Big Data from simple data is its bigger Volume, higher Variety and the rate at which it is collected. Collecting data from internet like reviews, recording sales in more details, etc. increases the amount of data. Volume and Variety of data regarding consumer, sales, inventory location, etc. is increased. Moreover frequency of data collection like hourly, daily, monthly, weekly or even with every click is also being increased. Sometimes data is collected in real-time also like with the help of sensors. Data is also becoming easier to store and access. Various softwares are available for data analysis. Manufacturers, service-providers, retailers, carriers, etc. can benefit by applying BD Analytics like quick response by manufacturer to customer demand, real-time capacity estimation by carriers, etc.

Many research questions arise about the SC Management applications in which BD Analytics be applied and how it can benefit in fields of Inventory management, customer and supplier relationship management, etc. Researchers have tried to answer these questions. Answering these questions has been the aim of the author for doing this project work.

INTRODUCTION

nRisk Management in Supply Chains is done using Big Data. Long and geographically wider Supply Chains and unavailability of to-the-point information make Supply Chains difficult to manage and give undesired results. Big Data measures taken company-wide can make Supply Chains more sustainable. Big Data can give a Supply Chain benefit over its competitors. According to Social Capital theory, companies of a Supply Chain should contribute while sharing information and analysing data. In the process they build strong relationships, accept norms of the Supply Chain well and this benefits in the competition. According to Resource Dependence theory, sometimes a company is depended on other companies for resources. Big Data can help in such situations like by keeping record of inventories it can remind to replenish it whenever they are less than safe value or some sudden surge in demand is expected. Resource Based View Theory suggests use of Big Data to increase productivity of resources like whenever a machine is not in use it should show being not in use automatically so that the next job waiting in queue for being machined can be machined on it thus increasing its productivity. Ecological Modernization theory suggests use of Big Data by Supply Chain management professionals to assess their environmental effects and remedies for the harmful like processing of health data in a locality can indicate rising level of pollutants thus suitable curing action can be taken.

Implementation of Big Data requires intra and inter companies systems. Information Technology systems are being gainfully employed. Well-designed systems can prove to be very efficient. The competition now-a-days is between Supply Chains instead of between individual firms. Large Volumes of data can prove harmful to existing data management systems of businesses also so proper system must be made to derive value from Big Data. Management of Supply Chains should include not only transportation facilities but also efficient production of goods and services. Thus Big Data applications involving multiple disciplines should be used. More and more utilization of data should be there. Big Data brings results like savings in maintenance bills, better quality, etc.

Knowledge workers are needed to run a Supply Chain Management system successfully and make it responsive to changing environment. A Knowledge Worker analyses the data and takes informed decisions. Information Technology facilities like email, social media, etc. and electronic gadgets aid SC professionals like Wikis are an online communication platform where professionals can share their views on an issue and reach a common solution.

Some researchers have tried to answer questions about Big Data through interview techniques. These interviews are based across the nations and many types of industries. The partners of a Supply Chain should share information freely in a Big Data system as they are going to from it in the long run.

Costs for implementing Big Data can be incurred in instalments duly assessing the returns. Big Data saves managers from taking decisions based on guesses which can be

Suitable professionals are required to implement Big Data in a firm. Those professionals should have knowledge about not only data-related techniques but also the field to which data is related.

What type of data is to be stored and for how much time is also an issue because storing and maintaining data is costly. Linking customers and suppliers to a company Big Data system can make it more profitable. Big Data can make a company procurement system more efficient by helping in judging trends, thus improving bargaining, quality, lead time, delivery and efficiency of the plant. Just-in- inventory systems are possible due to Big Data accuracy and trueness of data is must for it to be useful. Easy accessibility of data is also necessary for drawing benefits from Big Data. Records should be regularly updated.

Being a new concept, there is lack of guidance in implementing Big Data. So its implementation should be through learning. The system should be made user-friendl Big Data should be used to improve customer service by using it to make customized products offerings for customers. Future demands can be predicted thus profits can be increased. Skilled professionals are needed for deriving benefit from Big Data implementation. More research should be done in Big Data.

Collection and analysis of data can be done automatically also in Big Data like retail sector loyalty cards. It is claimed that companies which are employing Big Data can easily surpass their competitors. Taking initiatives and creativity can lead to high competitive advantage in Big Data. Electronic devices mounted on a ship can make maintenance work of its engines very efficient thus saving fuel, lubricants, etc. thus making tourism and distribution more sustainable. Theoretical different models should be made to implement Big Data indicating steps to be taken. What data is to be collected should be planned carefully. Real-time data collection can also be used. Sampling should be scheduled appropriately. Proper hardware for data acquisition and processing is needed Sensors, computers, personnel and presentation are

parts of a typical Big Data system. Online retailers like Amazon ship their products in advance by using prediction of Big Data Statistical methods and Probability theory can be used for making prediction Making simulation models and optimising them for best results helps in taking decisions. Sources of data can be many in a supply chain Conventional IT systems may not be able to handle Big Data. So new architecture has to be planned. Future risks can be predicted by Big Data. All types of industries are employing Big Data. Big Data can help avoid work stoppages due to lack of proper planning. Material handling work can be made easier by tagging products and handling units and Big Data technologies.

Modern cities can provide companies with facilities to implement Big Data, IT systems to support it . Big Data is used for judging pattern and correlations. Open databases can be used for many purposes different from those for which they were collected like AT CCTV data meant for customer safety can be used for assessing usage pattern of AT by the people Big Data helps in knowing consumption patterns thus helping in making distribution better. Big Data is especially useful in highly uncertain business situation

Big Data can be used to find solution to complex decision-making problems.

Big Data can be used to get insight into resource requirement and thus helps in planning. Studies conducted by many researchers across management of various companies through interviews, etc. has revealed that Big Data implementation helps in improving productivity of the facilities being used and growth of the business. Big Data can be employed across many functions of an organization specially marketing. Big Data may not give expected profits in early stages of implementation. Big Data helps in value generation in the Supply Chain. In any data system, we are storing data from many sources in a particular format and then retrieve information from it when it is required. Implementation of Big Data sometimes require integration across the supply chain. Indian Railway ticketing system, Google Search, an automated factory and Weather Forecasting stations are Big Data systems (real-time also) in our daily life. Models or simulation can be made using Big Data to understand and solve life problems. Big Data helps in selecting course of action from various alternatives available. Big Data like any other technology should be acceptable to employees of the company . Either skilled employees be recruited or internal employees should be trained. Any resistance shown by employees to change should be dealt with Big Data can help reduce returning of items in online shopping. Big Data can help integrate online and offline Supply Chains.

Risk Management in Supply Chains is needed for its smooth functioning. A business cannot be sustained if it is at risk. Big Data helps in taking better decisions by guiding on possible outcomes of various actions taken. Big Data helps in removing problems related to supplier. It helps in identifying and correcting defects in products and services. It helps in minimizing inventory. Big Data can improve delivery using weather forecasting, GP System, Maps, traffic updates. Accurate demand forecasting can be done by it using history, analytics, etc. Thus we can see there are a lot of advantages of using Big Data Analytics. Big Data can improve marketing and service after sales. Customized service can be given using Big Data. Faster service and response to issues arising in Supply Chain can be achieved by Big Data. Rise in product demand can be achieved by Big Data. Shorter delivery times and thus better customer and supplier relationships can be achieved using Big Data. One can get profits from data using Analytics tools.

Amazon analyzes every click on their site to predict demand and maintain stock at their stores accordingly. Loyalty cards data can be used to estimate consumption patterns for products, consumer choices, etc. In the company IBM, quality in the production is maintained using Big Data. The system warns about any quality issue at an early stage. Social media is used to get information of events which may disturb the Supply Chain

In DHL, a logistics company, improvement in delivery process, use of outside people for delivery is done using Big Data capacity of vehicle and warehouse is planned according to future requirement assessed through Big Data. Feedback from customer is taken action upon and service is improved continuously to keep customers loyalty are taken to tackle risks to the Supply Chain which are assessed by Big Data. Demand forecasting and preparation accordingly is done by Big Data. Real-time data collection from means like sensors on vehicles can be used for taking decisions for present and future both

Some companies are providing Big Data solutions for taking pricing decisions. Many issues are important for taking care of profitable operation of a Supply Chain. Implementing Big Data can provide remedies for such issues

A large milk and milk-product selling company maintains small and big collection centres and many vehicles. All activities have to be completed on time because milk is a fast-perishing product. Robberies, carelessness of staff, etc. issues are troubling management. For dealing with these issues the company has started using Big Data technologies. The

management is now able to reduce Supply Chain risks a lot. Vehicle Reports can be used for full utilization of fleet.

In a Delphi Study data can be collected from data sources both internal and external to companies. Big Data system can either improve already existing data-processing system or replace it. Big Data can be employed in Reverse Logistics also. In a Delphi Study to evaluate Big Data applications, officials from various companies were asked to participate.

In manufacturing processes, Big Data is used for knowing root causes issues, measuring performance, pricing, optimization, setting flexible production set-ups, automation, et

In planning, it helps in measuring consumption, determining supply strategy, assessing raw materials requirement, in assessing performance of planning, forecasting demand of new goods, emergency pl

In Logistics, it helps in planning the capacity, selection from various choices available, tracking products, developing network of stores, vehicles, et

In service,, it helps in knowing customer requirements, customizing services, designing new services, taking actions i in assessing which customers need direct marketing, taking actions as per customer feedback, differentiate between types of customers

In Sourcing it helps in rating supplier negotiation with suppliers on terms of supply, deciding upon which products to source making purchasing system intelligent

In Return, it helps in knowing root causes for defects, assessing Return process performance

In understanding what Big Data is, the study gives results that in managing Data it is procedure and use data, keeping well-connected data network

In collection of data it is various sources like product data, market data, supplier and purchaser data finance data, big volume, variety of data like, numbers, video, et great speed of data even real-time, automated collection like sensors, GP System

In using data it is determining optimum values from collected data, knowing issues and chances of profit, having insight, automatic decision-taking system like removal of a defective piece from the line by the robotic system which can measure quality

In managing data it is better processing of data, better data handling infrastructure, better integration of information network,

In performance, it is better quality, speed of producing goods/services, flexibility of the operations, lesser costs, meeting customer needs Quality

Inaccurate, incomplete, not updated and irrelevant data can lead to losses. Statistical control methods can be used to assess and maintain quality of large quantity of data

Benefits of Big Data

Customer-related

- savings can be achieved
- more elaborate services can be provided

How

- By never disappointing customer because of non-availability of goods
- Better processing of data

Product-related

- More profit due to new product, business method
- More profit due to expansion of existing business

How

- Designing new product and its release
- Removal of problems from existing product

Deriving value from Big Data should be the aim of the management. There are various IT technologies to do that

Data

Various research papers have been published on Big Data indicating its relevance in the present scenario

Sources of Big Data:-

Supplier, manufacturer

Storage

Hard-disk

Big Data helps in planning, making strategies, taking decisions and improving efficiency. Thus using data to add value to the business results in increasing profits. Big Data system emphasizes on integrating devices, machines in business to the Big Data network for collecting data. This data from other sources is then processed for adding value to the business.

In the present day scenario, SC Management has become a difficult task. More cooperation among its members is required to achieve goals. To remain in competition, a supply chain has to be efficient, responsive, flexible and result-oriented. Big Data technology can play a major role in achieving these characteristics.

In the present business scenario with fast-changing products, availability of different varieties of the same product to remain in competition SC Management has to be very responsive and opportunistic

An organization must be able to sense change and plan strategy accordingly

Processing of recorded data and real-time data is done by management to assess whether the performance of the Supply Chain is as per plan or not.

Big Data helps in building capability of Supply Chain to survive in dynamic environment

Big Data suggests ways for maximizing profits while working with constraints. For better advertisement and assessment of demand we should be able to acquire and process data in real-time like taking decisions according to insight.

Big Data focuses on prediction unlike traditional database systems. The type of processing depends on the type of data. Processing of structured data is done in a different way than that unstructured data

Sharing of data is necessary for an efficient Big-Data based system. The working of a Supply Chain depends on information, so applying Big Data techniques for SC Management can give good result

Companies are collecting large volume of data (in real-time) due to internet, electronic devices, etc. thus seen as a resource from which value should be derived.

The modern day technological advances also known as Industrial Revolution four, Production scenario centred at information which travels at high speed

Using past data, predictions about future demand, etc. can be made. Ma can be done on time thus avoiding hurries. Big Data helps in reducing complexities in the function of the Supply Chain

Risks are associated with use of low quality data. Big Data cannot replace traditional data management systems but should be used in association with them

Me data systems work on only internal information i.e. information which relates to a company or a Supply Chain but Big Data uses information from any source internal or external provided it is of high quality. Big Data derives value from that data also which doesn't have explicit information by means of its processing tools. The growing interest of researchers in Big Data indicates its relevance today. Many articles are being published on this topic every year. Big Data is a means to optimize Supply Chains. The businesses should be more data-based. More research can be done to figure out if Bi Data can be used for benefit. Professionals from various concerned fields like IT, Marketing, Production, Logistics, etc. are needed for making a result-giving big data system. Using social media posts to get ideas about future demand, predicting customer behaviour, develop new products, know trends, etc. is part of Big Data. Using RFID technology, suitable data processing softwares etc. one can make system which can take decisions to improve control of production. Using sensors,, et one can achieve optimization of transport systems. An organization using Big Data system should treat information as importantly as it treats products/services it produces. Supply Chain partners should share innovative ideas among themselves to become competitive in the marketing data can help in better production control like it can tell present location and availability of Material Handling vehicles on a large Shop Floor through the use of sensors. T can suggest improvement in strategies being followed for increase in profits. Big Data can help in finding out patterns in the data which are difficult to be perceived manually. To know the view of experts regarding Bi Data, Delphi Study was conducted by some researchers. Officials of both technical and managerial background were interviewed for the purpose. This was done to get a multi-perspective view. The expert panel was constituted taking care of type of qualification, level of qualification and experience in a

particular field of officials no further rounds of interview are conducted once we start getting fixed response. This is done also because experts may refuse to participate due to lack of time. Experts were given a time of two weeks. If someone didn't answer he was sent reminder to answer. If he still didn't answer he was removed from study. In the first round the respondents were allowed to answer liberally However in subsequent rounds(quantitative) they had to answer the questions(on Likert scale from 0 to 5) based on constructs made from answers of the first round. The results of a Delphi Study are generally considered very reliable. As per the study, on the corporate level, the field in which the expert think big data has maximum opportunity to deliver results are:-

Judging the behaviour of customer

- improve the visibility and transparency of the Supply Chain

On the supply chain level, the areas in which big data has opportunity are:-

- Transportation and related activities
- Improvement in transparency of the supply chain
- Increase in efficiency of operations
- Inventory

The challenges which are to be overcome by Big Data implementation

Corporate (management level):-

- Building IT infrastructure

On supply chain (working) level

- Personnel complying with the strategy
- Cooperation among supply chain partners

The constructs which get lower ranking doesn't mean are irrelevant but it means they have lesser importance when compared to higher ranking ones

According to the study, following can be the benefits of Big Data:-

- better visibility in the Supply Chain
- Real-time prediction leads to better maintenance and thus increased efficiency of equipment

- Better efficiency of operations
- Better response to outside like market, et fluctuation
- New models of business can be designed
- Increase in visibility of supply chain makes it more, flexible, dynamic
- Automated control based on data collected optimize the processes

Better integration leads to more trust-building among partners which increases efficiency

Freed up resource due to better utilization can be used for innovative activities

Challenges

High investment in IT infrastructure, upgradation can lead to wastage

Clear knowledge of why Big Data is required and what strategy to adopt to derive its full benefits

What type of skill-set is required in a Bi Data professional or existing personnel can handle it.

Participation of all stakeholders in big data initiative (compliance)

A good governance in the supply chain to direct members towards a common goal

Transparency in information sharing and at a higher speed in the supply chain leads to optimized inventory, lead times of planning cycle are reduced and planning can be done in a more detailed manner. Predictive maintenance is possible as early detection of problems can be done using advanced statistics methods like comparing real-time measured values with pre-set standards, use of better manufacturing planning , use of optimized transportation system like multimodal transport, etc. improves the efficiency of operations

For unhindered passage of information integration and cooperation in the supply chain is necessary. Information required by a company may not be available in the company itself use of intelligent manufacturing systems for checking and planning of production in a large area setup may help in this.

Security of data is also an important issue.

Some people don't trust Delphi Study results much as they think they are personal view an can't be generalize but they are of great value to the researcher

Experts from different industries and from different companies of the same industry should form the panel to get unbiased results

- Companies already strong in IT can easily implement Big Data
- Big Data can be a solution to several management working level problems
- Management in 1960's was just physical distribution but today it is also about cost-minimization, reliability, prediction, production optimization, et.
- Big Data helps in taking faster decision
- Big Data helps in improving quality of products and services
- Big Data enables SC professional to satisfy customers through various channels.

Big Data has potential to revolutionize SC Management. Use of Big Data has brought results in even Sport and election. Big Data helps in designing network of Supply Chain. Managing customer-related issues and using market opportunities well can be done by Big Data technologies. Business people should adopt Big Data to increase their output and competitiveness. Online retailers are judging customer requirement by Big Data tools and are easily surpassing their competitors who are not using Big Data tool. Customer transaction data, tag data internet mobile data etc. sum up to a huge data which cannot be processed by conventional data-processing tools but by Big Data tools. Now-a-days streaming data may also be processed. Business and IT leader should work together to derive value.

From data processing of data can be done in following ways

Descriptive

- Accessing database and show results
- Judging trends from it and presenting them

Predictive

- Forecasting future trend.
- Finding relationships between data which are not easily perceived

Evaluation of present strategies is done, new strategies which can be adopted are also determined keeping in mind goals to be achieved working within the constraints.

With the increasing size of Computer the quantity of data and the complexity of problem it can deal with also increase. Thus in the modern era of big-sized computers Big Data system are easy to implement. a

LITERATURE REIVEW

Empirical studies on Big Data in Supply Chain management

S. no.	Title of Research Paper	Authors	Year
1	A proposed Architecture for Big Data driven Supply Chain Analytic	Sanjib Biswas and Jaydip Sen	2016
2	Emerging Trends in Supply Chain Management	Penina Orenstein et al	2014
3	Big data Application in Operations/Supply Chain Management: A Literature Review	Richard Addo-Tenkorang and Petri Helo	2016
4	Big Data For Omni- channel Supply Chain Management	Matthew Roberts et al	
5	Big Data Analytics for Supply Chain Management: A Literature Review and Research Agenda	Samuel Fosso Wamba and Shahriar Akhter	2015
6	Data Science, Prediction and Big Data: A Revolution that will transform Supply Chain management	Matthew and Stanley	2013
7	Mitigating Supply Chain risk using Sustainability via Big Data	Mani et al.	2017

Ch. 1 BUSINESS DRIVERS FOR BIG DATA

Qualifying the business value is particularly important, especially when the forward-looking stakeholders in an organization need to effectively communicate the business value of embracing big data platforms, and correspondingly, big data analytics. For example, a justification might show how incorporating a new analytics framework can be a competitive differentiator.

1.2 UNDERSTANDING THE BUSINESS DRIVERS

By seeing some latest tendency for data production, utilization, sharing and the requirement for processing we get to know how gradually developing business scenario where implementation of big data is possible.

Business enhancers are about flexibility in use and processing of data collections and streams to generate value: profit increasing, Cost reduction, enhancement of customer service, lessening dangers, and rising productivity.

Big data requires taking, keeping accessing and processing data. Some main tendencies that enhance the requirement for big data systems include:-

- Large data sizes being recorded and stored: According to a study, “In 2011, the size of data generated and copied will be more than 1.8 zettabytes, . . . increasing by nine times in only 5 years.”³ The size of this increase is more than the normal capacity of conventional relational data handling systems, or even normally hardware designs supporting file dependent information retrieval.
- Fast speed of data rise: After only one year, study stated, “From 2005 to 2020, the digital data will increase by a 300 times, from 130 exabytes to 40,000 exabytes, or 40 trillion GB (more than 5,200 GBs for every man, woman, and child in 2020). From now till 2020, the digital data will be about two times in two years.”⁴
- Large data size entered in the network: According to CIA Forecast, by 2016, yearly worldwide internal data is predicted to be 1.3 zettabytes.⁵ This rise in internal data is due to the rise in quantity of smartphone and other Internet-enabled devices, the increasing group of Internet users, the rising Internet speed provided by telecommunications providers, and the

risen of wi-fi connectivity. Large data being sent into broader communication networks generate need for recording and handling that data in a timely way.

- Rise in variety of kinds of information assets for processing: In difference to conventional ways for according structured data, data scientists want to take benefit of unstructured data recorded from various sources. Some sources show very less structure (like Internet or phone call details), while some others are unstructured or formats (like social media content that combines text, photos, audio, and videos).

To derive value from this, organizations must improve their present structured data processing system to include processing of this streaming data also.

- New and unsynchronized ways for helping data delivery:

In a structured system, there are clear marking of the 32011 IDC Digital Universe Study: extracting value from chaos, different tasks for data acquiring, like bulk file sending through a disk and tape storage method, or through file sending protocol on Internet. Now-a-days, data transfer and publishing is full of unforeseen variations, with data received from various linked sources like sites, transaction records and even open content from government data and social networking sites such as Twitter. This generates needs for fast recording and processing while keeping updation in various data group.

- Requirement of real-time inclusion of a processing results: There are many people in increasing variety of position who are using processing results. The rise is prominent in firms where business systems are joined to fully include processing models to improve performance. Like, a retail firm can observe real-time transactions of thousands of Stock Units (SKUs) at many places and judge trending sales on minute basis. Giving these larger data groups to a group of various business peoples for processing gives insight and abilities that didn't exist before: it allows purchaser to observe trends to make profitable decisions, inventory managers to distribute shelf space more efficiently at the store, pricing people to change prices fast. The big data systems must process and give results in limited time.

1.3 LOWERING THE BARRIER TO ENTRY

Helping business heads to take benefits of Big Data system in new ways is difficult for most companies. And the increasing system of generated data has moved large-scale processing abilities out of reach of companies smaller than the largest.

Most of the “big data” technology is not new. Until recently processing capabilities could not be afforded by smaller companies due to lack of resources. However Big Data is not so costly and has larger processing abilities than earlier systems due to it having recent technologies.

As we will see, technology for business intelligence and processing have developed so much that different businesses can use capabilities that in the past were limited to the largest firms with large budgets.

The changes in the environment make big data processing beneficial to all types of firms, while the market conditions make it employable. The grouping of simple models for development, a wider set of data processing tools has effectively lowered the barrier to entry, enabling many organizations to develop and test out.

1.4 CONSIDERATIONS

Whilhoug says the conditions now-a-days are requiring big data implementation in businesses but it is still not an easy task.

CH. 2 BUSINESS PROBLEMS SUITED TO BIG DATA

- the rise of data volumes to be consumed;
- the desire to use both structured and unstructured data;
- lowered barrier to entry for enabling scalable high-performance processing;
- reducing operational costs by leveraging commodity hardware;
- simplified programming and execution model for scalable applications.

2.1 VALIDATING (AGAINST) THE HYPE: ORGANIZATIONAL FITNESS

Even as the excitement around big data analytics has increased, it remains a technology-driven activity.

2.2 THE PROMOTION OF THE VALUE OF BIG DATA

That being said, a thoughtful approach must differentiate between hype and reality, and one way to do this is to review the difference between what is being said about big data and what is being done with big data.

2.3 BIG DATA USE CASES

.An examination of list helps us to divide most of these applications into following groups:

- Business intelligence, queries, report-generation, searches including many applications of searching, filtering, faster aggregation for reporting, trend assessing, optimized search.
- better performance for generating data management works, data collection and sorting, data conversion and deleting duplicate data.
- Non-database applications, such as image processing, text processing in preparation for publishing, genome sequencing, protein sequencing and structure prediction, web crawling, and monitoring workflow processes.

- Data searching and processing applications which include social network

Data processing, face identification, profile comparison, other kinds of text processing, web searching, machine learning, data retrieval, recommendations based on processing of personal data, optimized advertising and behavior prediction.

The main processing capabilities of the Big Data system can be divided into next basic groups as

- Counting works used for large data that can be divided and distributed among computing and storage machines such as aggregation (counting and summation).
- Scanning works that can be divided into parallel works such as sorting, trend recognition, etc.
- Modelling capabilities to forecast
- Storing large data volumes while providing comparatively faster access.

Processing softwares can join these main functions in different manner.

2.4 TYPICAL QUALITIES OF BIG DATA USAGE

Many of the usages told by Hadoop employing professionals are not new. There are many common applications. The availability of a cheaper better performance computing system enables more users to make these applications, use them for larger works or speed up the making and result-giving time. This combined with review of the various types of other applications hints that of the cases told as Big Data successes, the Big Data is more appropriate for finding remedies to business problems that involve one or more of below criteria:

1. Data throttling: Business problem has solutions but on old hardware, the outcome of the solution is not optimum because of limitations of hardware in handling Big Data.
2. Computation-related throttling: There are solutions, but they have not been used because the required computational performance is not possible by traditional systems.

3. Large size of data: The analytical application combines a multitude of existing large datasets and data streams with high rates of data creation and delivery.

4. Considerable data variety: Data from various sources is different in structure and type of content and sometimes data is not even structured.

5. Benefits of parallel processing: The data processing can be done faster through parallel processing of independent data.

Businesses having above criteria should use Big Data technology for increasing the profits. The processing which is obstructed by old type of hardware, involve large size of data uses different types of data and can benefit from parallel processing should be done by Big Data technologies.

2.5 VALUE OF BIG DATA USE

Big Data adds value to the organization by:-

- Increasing business: like by using a recommendation engine sales can be increased.
- Decreasing costs: like by better prod planning and inventory management cost incurred can be reduced.
- Increasing productivity
- Reducing risk

s

2.6 FORWARD THINKING ABOUT VALUE

While we continue employing big data technologies for developing solutions, more ways can be devised to take benefits from them. An organization must be ready for implementing Big Data, developing, and implementing big data applications.

CH. PREPARING ORGANIZATION FOR BIG DATA

The common use of the word “big data” may create a false picture in one’s mind that it is technically simple and this can be troubling for professional who want to use it. In fact, there are many aspects of big data technology, with their own technical details: hardware devices for processing, software that drive hardware, NoSQL and graph-using data managing systems, processin applications alongwith various applications included in Apache’s Hadoop software. A

.1 IMPORTANT QUESTIONS

s mentioned in earlier discussion, market situations and preparedness of the organizations have enabled professionals of an organization to check a various big data tools and technologies. Yet before any of these recent technologies can be employed in Supply Chain management system, they need to win confidence in wider organization setups that they caasetting in ways that add m cosnsiderable vganisation to become aware of a champion, comfortable business sponsorship, and establish bidirectional engagement to preserve the commercial enterprise sponsors knowledgeable and aware of progress.

A top example is the desire to circulate live information immediately into a large information analytical software for real-time integration, while data warehouses are often simply populated with static datasets extracted from present front-quit systems.

- present infrastructure: A decade (or extra) of investment inside the traditional statistics warehouse/enterprise intelligence framework has institutionalized positive processes to statistics control, yet the decision approximately the prevailing infrastructure (along with the traditional approaches to extracting facts from assets to load into the warehouse rather than newer processes to federation and virtualization that permit the records to remain in its authentic source) affects get entry to synchronized datasets in addition to usability of a ramification of statistics assets which can be predicted to feed huge information analytics.
- present funding: The word referenced in the preceding bullet bears repeating. companies have sunk tremendous investments in growing a BI/DW surroundings, and some attention

have to be applied to evaluate the diploma to which one wants to both salvage or abandon that investment.

- records purpose: maximum data instances are created for precise purposes, but big statistics applications seek to repurpose data for evaluation. The authentic information intent can also differ substantially from an array of potential uses, and this implies the need for extra governance for information control, nice, and semantic consistency.

- length and length: The desire to acquire and use big datasets has direct implications for the “pure” components of information control. The transitory characteristics associated with fast turnaround of numerous facts streams conflicts with the desire to preserve very big datasets in anticipation of the potential for brand spanking new analyses to be executed in the destiny. This tension will force organisations to make funding and capital acquisition selections to assist information staying power and retention.

3.2 RELATED TO THE PROPER SELECTION MAKERS

Given those demanding situations, how can corporations plan to support big facts? extra to the point: Who in the enterprise needs to be involved within the procedure of obtaining, proving, and then deploying massive facts answers and what are their roles and responsibilities?

In any era adoption cycle, it's miles incumbent upon the important thing stakeholders inside the employer to make certain that the business procedure proprietors, the information clients, the technical infrastructure innovators, the software developers, and the employer architects all paintings collectively in an surroundings that may keep to satisfy current reporting needs yet is bendy enough for exploratory paintings.

We will look at a widespread sequence of responsibilities to help us remember the way to control the transition right into a manufacturing development manner to take gain of the enterprise fee large facts strategies can offer. The sequence begins with recognizing this opportunity, then defining expectancies, and piloting, vetting, and assessing big information technology earlier than moving into production.

r1. apprehend the opportunity: Use the standards in bankruptcy to identify the commercial enterprise demanding situations which have the most to benefit from adopting one or extra of the massive statistics techniques.

2. define expectancies: Articulate specific value development expectations inside a bounded time frame. those expectancies are the basis for a later pass/no-pass choice.
3. Scope a pilot: Spec out the framework for what ideas are to be verified throughout the development and assessment of a pilot application.
4. Collect and manage assets: pick out and gather the talents essential for huge statistics improvement.
5. Application control: Devise a application plan for the pilot phase in addition to expect the tasks for migration into production.
6. Improvement and assessment: design, broaden, code, take a look at, and examine whether the end result meets expectancies.
7. Flow into production: If the technology appears to add cost, make stronger the plan to transport it into production.

Soliciting clear business requirements and specifying corroborating success standards and measures permits a level of consider that test-driving new generation like big information analytics is not simply an highbrow workout, however rather is a repeatable procedure for injecting new thoughts into the agency.

4Integration of big records technology into an corporation as with any new and better designed technology that may growth enterprise value, many professionals have adopted big records processing as a supporting way, if now not major means of reporting and facts-processing for the employer. And as with the employment of any new technology, there is a need to actual that even though the organisation is prepared to adopt the era problems might also emerge as obvious at a later degree.

It is unwise to use a new technology without determining its ability cost to improve the present methods or generate new opportunities. This fee is shown in shape of increased earnings, reduced risks, higher consumer enjoy, multiplied business, and so forth. which need to be more than the costs of introducing the generation and its non-stop operation.

Testing and piloting technology is necessary to preserve an company's competitiveness and assessing that the new era is appropriate for use. There are measures for success and acceptability which are really described in the context of these dimensions of business price.

However in many agencies, the methods to apply new technology and tools frequently bypass practices which are crucial for easy working of latest technology with existing ones. Pilot tasks are released to check the new technology with out proper planning with the end result that pilot projects that are hurriedly moved into testing provide no evaluation of the new era regardin its performance or architecturive.

Ch 4 USE CASES

4.1 CONSIDERING PLATFORM ALTERNATIVES

While considering the special methods of deploying an analytics surroundings, the key selections for investing in infrastructure attention on how the platform high-quality meets the expected performance desires. One need to be willing to specify key measures for machine performance to correctly investigates capability requirements for the meant analytical packages to help pick out a selected architectural method.

The benefits of the use of hardware appliances for huge data center on engineering and integration. they're engineered for high-performance reporting and analytics, but have a flexible architecture allowing incorporated components to be configured to satisfy particular software desires.

In a column-oriented database, simplest the columns inside the query want to be retrieved. And while there is a capital funding in equipment, hardware appliances are low fee when in comparison to big information warehouse hardware structures.

One advantage of the use of software appliances, in the meantime, is that they can take gain of low-fee commodity hardware additives. In addition, the reliance on commodity hardware allows a software equipment to be elastic and extensible.

however, you need to keep in mind all factors of the performance desires of the one-of-a-kind sorts of applications: facts scalability, person scalability, get right of entry to and loading pace, the want for workload isolation, reliance on parallelization and optimization, reliability within the presence of disasters, the dependence on garage duplication or information distribution and replication, among other overall performance expectations. Then take a look at how the performance needs of the one of a kind types of packages are addressed through each of the architectures. this could provide a measurable method for assessing technology suitability.

Ch. 5 BIG DATA EQUIPMENT AND TECHNIQUES

5.1 ABOUT BIG DATA STORAGE

As we've got mentioned in lots of the report to this point, maximum, if no longer all massive records programs attain their performance and scalability through deployment on a collection of garage and computing assets certain together within a runtime environment. In essence, the ability to design, expand, and enforce a big records application is directly dependent on an awareness of the architecture of the underlying computing platform, both from a hardware and greater importantly from a software program angle.

One commonality among the unique appliances and frameworks is the variation of equipment to leverage the combination of collections of key computing resources:

1. Processing functionality, regularly referred to as a CPU, processor, or node. usually talking, cutting-edge processing nodes often include more than one cores which might be character CPUs that proportion the node's memory and are controlled and scheduled together, allowing a couple of duties to be run simultaneously; that is called multithreading.
2. Reminiscence, which holds the information that the processing node is currently working on. most single node machines have a restriction to the quantity of reminiscence.
3. Storage, providing endurance of data—the area wherein datasets are loaded, and from which the facts is loaded into reminiscence to be processed.
4. Network, which presents the “pipes” via which datasets are exchanged among one of a kind processing and garage nodes.

Due to the fact unmarried-node computer systems are limited of their capacity, they cannot effortlessly accommodate massive quantities of facts. this is why the high-performance systems are composed of collections of computers in which the huge quantities of statistics and necessities for processing can be distributed among a pool of assets.

5.2 A WIDESPREAD REVIEW OF LARGE PERFORMANCE ARCHITECTURE

maximum high-performance platforms are created by means of connecting a couple of nodes together thru a ramification of network topologies. distinctiveness home equipment may also differ in the specifics of the configurations, as do software program home equipment. however, the overall structure distinguishes the control of computing assets (and

corresponding allocation of responsibilities) and the management of the records across the network of storage nodes.

In this configuration, a grasp job supervisor oversees the pool of processing nodes, assigns obligations, and monitors the hobby. On the equal time, a storage supervisor oversees the information garage pool and distributes datasets across the gathering of garage sources. Even as there's no a priori requirement that there be any colocation of records and processing responsibilities, it is useful from a performance angle to ensure that the threads technique facts that is local, or near limit the expenses of records get entry to latency.

To get a higher know-how of the layering and interactions within a large information platform, we are able to have a look at the Apache Hadoop software program stack, for the reason that architecture is posted and open for overview.

Hadoop is essentially a group of open source projects that are bodily combined to enable a software program-based totally massive statistics appliance. We begin with the core aspects of Hadoop's utilities, upon which the subsequent layer inside the stack is propped, specifically Hadoop distributed record systems (HDFS) and MapReduce. a new technology framework for process scheduling and cluster control is being evolved underneath the call YARN.

5.3 HDFS

HDFS tries to permit the garage of massive documents, and does this by using dispensing the information among a pool of facts nodes. A single name node (now and again referred to as NameNode) runs in a cluster, related with one or more records nodes, and provide the control of a standard hierarchical file employer and namespace. The call node efficiently coordinates the interplay with the distributed facts nodes.

The creation of a file in HDFS appears to be a unmarried report, despite the fact that it blocks "chunks" of the document into pieces which might be saved on individual data nodes.

The name node continues metadata about every record as well as the history of adjustments to record metadata. That metadata consists of an enumeration of the controlled documents, properties of the documents, and the document

gadget, in addition to the mapping of blocks to documents at the information nodes.

The records node itself does not control any statistics about the logical HDFS report; as an alternative, it treats each information block as a separate document and shares the vital statistics with the name node.

As soon as a file is created, as information is written to the document, it is truly cached in a transient document. whilst the quantity of the facts in that temporary document is sufficient to fill a block in an HDFS record, the name node is alerted to transition that temporary report right into a block this is committed to a everlasting statistics node, which is also then included into the file management scheme.

HDFS presents a degree of fault tolerance via information replication. An application can specify the diploma of replication (i.e., the range of copies made) whilst a report is created. The name node also manages replication, attempting to optimize the marshaling and conversation of replicated records in terms of the cluster's configuration and corresponding efficient use of network bandwidth. this is increasingly critical in large environments including multiple racks of records servers, for the reason that communication among nodes on the equal rack is normally quicker than between server node sin exclusive racks.

HDFS attempts to keep attention of records node locations across the hierarchical configuration.

In essence, HDFS offers overall performance thru distribution of information and fault tolerance through replication. The end result is a stage of robustness for reliable massive file garage. allowing this stage of reliability must be facilitated through a number of key responsibilities for failure

management, some of which are already deployed inside HDFS at the same time as others are not presently carried out:

- monitoring: there may be a non-stop "heartbeat" conversation among the facts nodes to the call node. If a information node's heartbeat is no longer heard by using the call node, the records node is considered to have failed and is not to be had. In this situation, a replica is hired to replace the failed node, and a trade is made to the replication scheme.
- Rebalancing: that is a procedure of automatically migrating blocks of information from one records node to some other when there is free area, while there may be an increased demand for the statistics and shifting it is able to improve overall performance (such as shifting from a

conventional disk drive to a stable-nation drive that is much faster or can accommodate extended numbers of simultaneous accesses), or an multiplied want to replication in response to greater common node screw ups.

- managing integrity: HDFS uses checksums, which can be effectively “digital signatures” associated with the real information stored in a file (regularly calculated as a numerical feature of the values within the bits of the documents) that may be used to confirm that the statistics stored corresponds to the information shared or obtained. while the checksum calculated for a retrieved block does not identical the saved checksum of that block, it's far considered an integrity blunders. if so, the requested block will want to be retrieved from a duplicate rather.

- Metadata replication: The metadata documents are also situation to failure, and HDFS can be configured to maintain replicas of the corresponding metadata documents to shield towards corruption.

- Snapshots: that is incremental copying of facts to establish a point in time to which the machine may be rolled returned.¹

¹Some very good facts is to be had in these ideas map to unique internal protocols and offerings that HDFS makes use of to permit a huge-scale statistics management report system that can run on commodity hardware additives. The ability to apply HDFS solely as a method for creating a scalable and expandable file system for preserving speedy get entry to to huge datasets provides a affordable value proposition from an facts generation angle:

- decreasing the fee of specialty big-scale storage systems;
- providing the ability to rely on commodity additives;
- enabling the ability to installation the usage of cloud-primarily based services;
- reducing device management costs.

5.4 MAPREDUCE AND YARN

at the same time as MapReduce is mentioned in greater detail in chapter eight, it's far precious to each introduce the overall concept of activity manage and management.

In Hadoop, MapReduce at the start mixed each process control and oversight and the programming version for execution.

The MapReduce execution surroundings employs a master/slave execution version, in which one grasp node (referred to as the JobTracker) manages a pool of slave computing resources (known as TaskTrackers) which might be known as upon to do the real paintings.

The function of the JobTracker is to control the assets with a few unique responsibilities, consisting of managing the TaskTrackers, always monitoring their accessibility and availability, and the one-of-a-kind

components of activity control that consist of scheduling tasks, monitoring the development of assigned duties, reacting to recognized disasters, and making sure fault tolerance of the execution. The role of the TaskTracker is a great deal simpler: await a undertaking undertaking, initiate and execute the requested undertaking, and provide fame lower back to the JobTracker on a periodic basis.

Exclusive clients can make requests from the JobTracker, which becomes the sole arbitrator for allocation of resources.

There are barriers inside this current MapReduce model. First, the programming paradigm is well desirable to packages in which there is locality among the processing and the data, but packages that call for records movement will unexpectedly grow to be bogged down by network latency troubles. second, not all packages are without difficulty mapped to the MapReduce version, but programs evolved the usage of opportunity huge records equipment and strategies sixty five programming strategies might nonetheless want the MapReduce system for process control. third, the allocation of processing nodes inside the cluster is fixed through allocation of certain nodes as “map slots” as opposed to “reduce slots.” when the computation is weighted towards one of the levels, the nodes assigned to the other phase are in large part unused, resulting in processor underutilization.

This is being addressed in future versions of Hadoop via the segregation of responsibilities inside a revision called YARN. in this approach, overall useful resource control has been centralized whilst management of sources at every node is now executed via a local NodeManager.

Similarly, there may be the idea of an ApplicationMaster this is related with every utility that at once negotiates with the central ResourceManager for assets whilst taking up the responsibility for monitoring development and monitoring reputation. Pushing this responsibility to the utility environment lets in more flexibility within the assignment of resources as well as be extra effective in scheduling to improve node utilization.

Remaining, the YARN technique permits programs to be better aware about the records allocation throughout the topology of the assets inside a cluster.

This cognizance permits for stepped forward colocation of compute and facts assets, decreasing information motion, and therefore, lowering delays associated with statistics access latencies. The result must be increased scalability and overall performance.²

5.5 INCREASING THE BIG DATA SOFTWARE ENVIRONMENT

At this factor, some key factors concerning the improvement of massive records applications must be clarified. First, notwithstanding the simplicity of downloading and installing the center components of a huge statistics development and execution environment like Hadoop, designing, developing, and deploying analytic applications nonetheless calls for a few skill and understanding. Second, one should differentiate between the duties associated with application layout and improvement and the responsibilities associated with architecting the massive statistics gadget, choosing and connecting its components, system configuration, in addition to machine monitoring and endured protection.

²An extraordinary description can be found in a sequence beginning with Murthy A. In different phrases, transitioning from an experimental “laboratory” device right into a manufacturing environment needs extra than simply get right of entry to to the computing, reminiscence, storage, and network sources. there may be a want to amplify the atmosphere to comprise a diffusion of additiona abilities, including configuration management, statistics agency, utility development, and optimization, in addition to extra abilities to aid analytical processing. Our examination of a prototypical massive facts platform engineered the use of Hadoop maintains by way of searching at some of additional additives that might usually be considered as part of the ecosystem.

5.6 ZOOKEEPER

each time there are more than one obligations and jobs walking within a unmarried allotted surroundings, there may be a want for configuration management and synchronization of various components of naming and coordination.

The venture's internet web page specifies it more honestly: "Zookeeper is a centralized provider for preserving configuration statistics, naming, presenting distributed synchronization, and supplying group services."three

Zookeeper manages a naming registry and effectively implements a gadget for dealing with the numerous static and ephemeral named items in a hierarchical manner, just like a report device. similarly, it enables coordination for exercising manage over shared sources that are impacted with the aid of race situations (in which the anticipated output of a manner is impacted via variations in timing) and deadlock (in which multiple obligations vying for manage of the identical resource correctly lock every different out of any venture's capacity to use the resource). Shared coordination offerings like those provided in Zookeeper allow developers to hire these controls without having to expand them from scratch.

5.7 HBASE

HBase is any other example of a nonrelational information management surroundings that distributes big datasets over the underlying Hadoop framework. HBase is derived from Google's BigTable and is a column-orientated statistics layout that, when layered on pinnacle of Hadoop, affords a fault-tolerant approach for storing and manipulating huge massive records equipment and techniques. As turned into mentioned in chapter 6, information stored in a columnar layout is amenable to compression, which will increase the amount of facts that may be represented whilst decreasing the real storage footprint. In addition, HBase supports in-memory execution.

HBase isn't always a relational database, and it does not aid square queries. There are a few simple operations for HBase: Get (which access a specific row inside the desk), positioned (which shops or updates a row inside the table), scan (which iterates over a group of rows inside the desk), and Delete (which removes a row from the desk). due to the fact it could be used to arrange datasets, coupled with the performance furnished with the aid of the factors of the columnar orientation, HBase is a reasonable alternative as a persistent garage paradigm whilst strolling MapReduce programs.

5.8 HIVE

One of the regularly-cited troubles with MapReduce is that although it affords a method for growing and executing applications that use massive amounts of information, it isn't more than that. And while the statistics may be controlled within documents the use of HDFS, many business programs assume representations of records in dependent database tables. That became the inducement for the development of Hive, which (according to the Apache Hive net site⁴) is a “facts warehouse machine for Hadoop that helps smooth information summarization, ad-hoc queries, and the evaluation of large datasets saved in Hadoop like minded file systems.” Hive is in particular engineered for information warehouse querying and reporting and is not meant to be used as inside transaction processing systems that require real-time question execution or transaction semantics for consistency on the row level.

Hive is layered on pinnacle of the record device and execution framework for Hadoop and permits packages and customers to organize statistics in a structured facts warehouse and consequently query the records the usage of a question language referred to as HiveQL that is similar to square (the usual established query Language used for maximum present day relational database control systems). The Hive machine offers gear for extracting/reworking/loading information (ETL) into a diffusion of various facts codecs.

And due to the fact the records warehouse device is constructed on top of Hadoop, it permits native get entry to to the MapReduce model, permitting programmers to develop custom Map and decrease features that can be at once integrated into HiveQL queries. Hive provides scalability and extensibility for batch-style queries for reporting over big datasets which might be normally being elevated while counting on the faulttolerant aspects of the underlying Hadoop execution version.

5.9 PIG

even though the MapReduce programming version is especially straightforward, it still takes a few skill and know-how of both parallel and allotted programming and Java to pleasant take gain of the version. The Pig mission is an strive at simplifying the utility development method with the aid of abstracting a number of the info away via a better stage programming language known as Pig Latin. consistent with the project's internet site⁵, Pig's excessive-level programming language permits the developer to specify how the evaluation is completed. In flip, a compiler transforms the Pig Latin specification into MapReduce packages.

The reason is to embed a full-size set of parallel operators and features contained inside a manipulate collection of directives to be implemented to datasets in a manner this is fairly similar to the manner sq.statements are carried out to standard structured databases. a few examples include generating datasets, filtering out subsets, joins, splitting datasets, removing duplicates. For easy applications, the use of Pig offers huge ease of improvement, and more complicated obligations can be engineered as sequences of applied operators.

imilarly, the use of a high-degree language also allows the compiler to perceive opportunities for optimization that would had been unnoticed by means of an inexperienced programmer. on the equal time, the Pi environment permits developers to create new user described functions (UDFs) that can subsequently be incorporated into advanced packages.

5.10 MAHOUT

attempting to use large statistics for analytics might be limited with none analytics skills. Mahout is a challenge to provide a library of scalable implementations of machine learning algorithms on pinnacle of MapReduce and Hadoop. As is defined at the assignment's home page6, Mahout's library includes severa well-known evaluation methods along with:

- Collaborative filtering and different consumer and item-based totally recommender algorithms, that's used to make predictions about an person's hobby or options through assessment with a multitude of others which could or might not percentage similar traits.
- Clustering, inclusive of okay-method, Fuzzy ok-way, suggest Shift, and Dirichlet manner clustering algorithms to look for companies, styles, and commonality amongst decided on cohorts in a population.
- Categorization using Naïve Bayes or choice forests to area objects into already defined categories.
- textual content mining and subject matter modeling algorithms for scanning textual content and assigning contextual meanings.
- common sample mining, which is used for marketplace basket analysis, comparative health analytics, and different patterns of correlation inside massive datasets.

Mahout also helps other methods and algorithms. the provision of implemented libraries for these sorts of analytics free the development crew to take into account the forms of problems to be analyzed and greater specifically, the kinds of analytical models that can be carried out to are trying to find the nice solutions.

5.11 CONCERNs

big records analytics packages employ a ramification of tools and techniques for implementation. while organizing your mind about growing those applications, it's far crucial to reflect on consideration on the parameters to be able to body your needs for technology assessment and acquisition, sizing and configuration, methods of statistics organisation, and required algorithms for use or evolved from scratch.

previous to diving directly into downloading and putting in software, attention on the types of massive statistics commercial enterprise packages and their corresponding performance scaling desires, consisting of the ones listed in table 7.1.

The technical requirements will manual each the hardware and the software configurations. This additionally lets in you to align the development

of the platform with the enterprise utility development desires.

7

S

C CH.6 C++ PROGRAM

Below is a C++ program demonstrating how processing by modern means of collected data can lead to maximization of profits.

Problem statement:-

A manufacturing company is engaged in producing three types of products: A, B and C. The production department produces, each day, components sufficient to make 50 units of A, 25 units of B and 30 units of C. The management is confronted with the problem of optimizing the daily production of the products in the assembly department, where only 100 man-hours are available daily for assembling the products. The following additional information is available:

Type of Product	Profit Contribution per Unit of Product (Rs.)	Assembly Time per Product (hrs.)
A	12	0.8
B	20	1.7
C	45	2.5

The company has a daily order commitment for 20 units of product A and a total of 15 units of products B and C.

LP model formulation

Decision variables Let x_1 , x_2 and x_3 = number of units of products A, B and C to be produced, respectively.

The Linear Programming model

Maximize (total profit) $Z = 12x_1 + 20x_2 + 45x_3$

subject to the constraints

(i) Labour and materials

(a) $0.8x_1 + 1.7x_2 + 2.5x_3 \leq 100$; (b) $x_1 \leq 50$; (c) $x_2 \leq 25$; (d) $x_3 \leq 30$

(ii) Order commitment

(a) $x_1 \geq 20$; (b) $x_2 + x_3 \geq 15$

and $x_1, x_2, x_3 \geq 0$.

By using the following C++ program the above LP problem was solved giving the solution as $x_1=31, x_2=0, x_3=30$ and the maximum profit as Rs. 1722.

```
#include<iostream.h>
#include<conio.h>
void main();
{
clrscr();
int l, o[4];
float s[3][4], e[3];
int j, k=3;
for (j=0;j<3;++j)
{
cout<<" enter maxi. Function coef."<<j<<"-";
cin >>e[j];
}
for(j=0;j<k;++j)
{
for(l=0;l<=3;++l)
{
cout<<"constraint equation"<<j<<" coefficient"<<l<<"-";
cin>>s[j][l];
```

```

}
cout<<"constraint equation"<<j<<"operator (1:>, 2: >=, 3: <, 4: <-";
cin>>o[j];
}
float m, n, p, m1, n1, p1;
cout<<"maxi. 1=";
cin>>m1;
cout<<"maxi. 2=";
cin>>n1;
cout<<"maxi. 3=";
cin>>p1;
float b, nw=0;
float so[3];
int solu;
for(m=0;m<=m1;++m)
{
for(n=0;n<=n1;++n)
{
for(p=0;p<=p1;++p)
{
solu=1;
for(j=0;j<k;++j)
{
b=0;
for(l=0;l<3;++l)
{
switch(l)

```

```

{
case0: b=b+s[j][0]*m;
        break;
case1: b=b+s[j][1]*n;
        break;
case2: b=b+s[j][2]*p;
        break;
}
}
switch(o[j])
{
case1:if(!(b>s[j][3]))
        solu=0;
        break;
case2: if(!(b>s[j][3]||b==s[j][3]))
        solu=0;
        break;
case3:if(!(b<s[j][3]))
        solu=0;
        brea
case4:if(!(b<s[j][3]||b==s[j][3]))
        solu=0;
        break;
}
}
if(solu==1)
{

```

```
b=e[0]*m+e[1]*n+e[3]*
```

```
If(b>nw)
```

```
{
```

```
so[0]=m;
```

```
so[1]=n;
```

```
s
```

```
nw=b;
```

```
}
```

```
}
```

```
}
```

```
}
```

```
}
```

```
cout<<"\n\nThe solution is - "<<so[0]<<" "<<so[1]<<" "<<so[2]<<" maxi. Profit= "<<nw;
```

```
getche();
```

```
}
```

The solution screen is:-

```
enter maxi. function coef.0-12
enter maxi. function coef.1-20
enter maxi. function coef.2-45
constraint equation 0 coefficient 0-.8
constraint equation 0 coefficient 1-1.7
constraint equation 0 coefficient 2-2.5
constraint equation 0 coefficient 3-100
constraint equation 0 operator(1:>,2:>=,3:<,4:<=- 4
constraint equation 1 coefficient 0-1
constraint equation 1 coefficient 1-0
constraint equation 1 coefficient 2-0
constraint equation 1 coefficient 3-20
constraint equation 1 operator(1:>,2:>=,3:<,4:<=- 2
constraint equation 2 coefficient 0-0
constraint equation 2 coefficient 1-1
constraint equation 2 coefficient 2-1
constraint equation 2 coefficient 3-15
constraint equation 2 operator(1:>,2:>=,3:<,4:<=- 2
maxi. 1=50
maxi. 2=25
maxi. 3=30
```

The solution is- 31 0 30 1722

H.6

CONCLUSION AND DISCUSSION

The research gap identified in the previous semester in the field of “Role of Big Data in Supply Chain management” was that a thorough study was lacking regarding how Big Data can benefit Supply Chain management professionals. In the author’s viewpoint he has succeeded in elaborating this in his research work as explained in this report. The concept of Big Data is not very old and everyday knowledge about its usefulness is increasing with more and more people and businesses employing it. It is not just limited to devices but also the ways in which the huge data collected can be utilized. It can be surely said that Big Data will keep on making our lives simple.

The author is also grateful to people whose work he has referred in doing this research.

REFERENCES:-

1. Sanjib Biswas et al, "A proposed Big Data architecture"
2. Penina Orenstein et al, 2014, "Emerging Trends in Supply Chain Management"
3. Richard Addo-Tenkorang and Petri Helo,2016, "Big data Application in Operations/Supply Chain Management: A Literature Review"
4. Matthew Roberts et al, "Big Data For Omni- channel Supply Chain Management"
5. Samuel Fosso Wamba and Shahriar Akhter, 2015, "Big Data Analytics for Supply Chain Management: A Literature Review and Research Agenda"
6. Matthew and Stanley, 2013, "Data Science, Prediction and Big Data: A Revolution that will transform Supply Chain management"
7. Mani et al, 2017, "Mitigating Supply Chain risk via Sustainability through Big Data"