DATA GEEK

SCHOOL OF BUSINESS AND MANAGEMENT

BIG DATA

BY
BUSINESS ANALYTICS
SPECIALIZATION

EDITORS DESK

"Without big data analytics, companies are blind and deaf, wandering out onto the web like deer on a freeway." – Geoffrey Moore, Author and Consultant

The concept of big data has been around for years; most firms now recognize that if they capture all of the data that flows into their companies, they can apply analytics and derive tremendous value from it. Businesses were employing basic analytics to identify insights and trends as early as the 1950s, decades before the term "big data" was coined.

The new advantages of big data analytics, on the other hand, are speed and efficiency. Previously, a firm would have gathered information, run analytics, and discovered information that could be used for future decisions; now, that organization may identify insights for current business decisions.

With this, we present an enthralling Volume 3 Issue 3 of DataGeek newsletter, centered on a pivotal theme of Big Data Analytics. It includes interesting articles, white papers, puzzles, surveys with dashboards, infographics and projects done on Big Data Analytics by the students of BA specialization.

I would like to extend gratitude to our Dean, Dr. Jain Mathew, Associate Deans Dr. Georgy Kurien and Dr. Jeevananda S, Head of Specialization – Business Analytics, Dr. Lakshmi Shankar Iyer for their guidance in making this issue a success. Also, a special appreciation to the newsletter team for the effort, time and inputs without which this issue would not have been possible. A thanks to all the students who have provided their valuable contributions. Once again congratulations to the entire team.

Please reach out to us for any queries or suggestions at datageek@mba.christuniversity.in

WITH REGARDS,
DR. TRIPTI MAHARA

CONTENT

INTELLIGENT TRANSPORT SYSTEM

INTRODUCTION

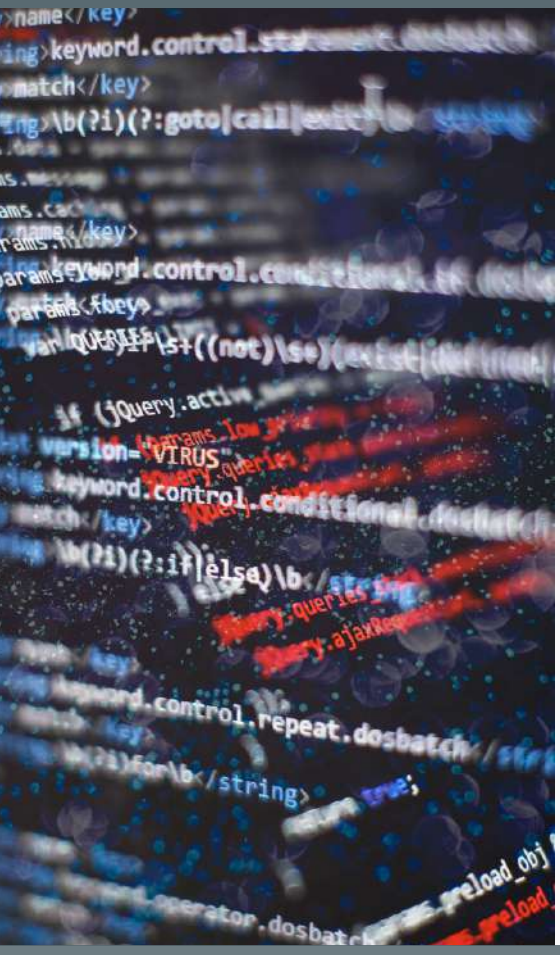
With the promotion of city development level in our country, traffic jams have been more and more serious. Facing the current traffic problems, intelligent transportation comes into being. An intelligent transportation system integrates electric sensor technology and information technology, which can help to enhance the quality of transportation management to some degree. But it is the premise to achieve intelligent transportation construction that accurately and timely acquires varieties of traffic data and constructs the handling model of traffic data through big data technology.

BUSINESS PROBLEM

Transportation allows a country to access natural resources and stimulate trade, allowing it to build wealth and power. Due to the increase in the population, the control over traffic management has become one of the challenges to the traffic department. This led to the rise in public transportation, especially government buses. Lack of knowledge in estimating the demand in the market on a daily basis vehicle operations has become inefficient. With no proper planning and scheduling running of buses leads to the inappropriate use of both manpower and fuel. The excess workload on the employees and lack of required infrastructure and communication with the vehicles worsen the safety of the employees and commuters. Customer service is one department that increases customer satisfaction and customer base for any organization, especially for the transport department. The lack of communication between the customers and service providers is creating the movement of customers to private transportation with a decline in the organization's profits.

"Intelligent Transport System"

"Intelligent Transport System"



ARCHITECTURE

In this intelligent transport system, ETL is an architecture (extract transform, load of data) used. ETL architecture was developed under apache-spark, which is used to process large data, including cleaning and transformation. The architecture addresses the following technical requirements

- It should be able to deal with raw data in many formats and sizes with the data quality
- Efficient big data transformation and storage to process data from traffic sensors.

To address this, apache-spark is needed. Spark is a framework used to store and process the data in real-time across various clusters. We input the data like CSV files, including tolling data, and unstructured data like images and videos. SparkSQL provides a framework for data management for different types of data like images, videos. This then stored in MongoDB. MongoDB can be used as a NoSQL approach for storing and managing the traffic data because it is scalable and easily integrated with the spark. MongoDB was configured by having two types of data, one is of real-time data: sensor data, traffic events (accidents, road works), and the other is historical data.

RESULTS/SOLUTION

- Allocation of public commute to the areas of maximum demand to avoid overcrowding.
- An integrated transportation system helps reduce oil imports for the government of India.
- It connected public transportation app for better knowledge of timing and whereabouts of public commute.
- Use of block chain technology for better security and decentralized model.

CONCLUSION

The construction of intelligent transportation systems is an inevitable trend of modern transportation industry development, and it also is the revolutionary transition of the traffic industry. And the application of big data technology provides the possibility for the achievement of intelligent transportation, and it further enhances the automation, integration, and information level of intelligent transportation. For example, traffic guidance, traffic transportation security, public transportation service, and active, intelligent transportation service, which fully proved the application possibility and necessity of big data technology in such industries. In the future, big data technology in intelligent transportation will be bound to be involved in big data handling technology.

Venkata Sai Krishna Nivarthi
2028232

Choragudi Vamsi Krishna
2027829

UNIQUE IDENTITY PROJECT

"The Unique Identity Project"

PROBLEM STATEMENT

The main agenda of Digital Ecosystem is assigning a unique identity to every Indian citizen. This agenda is achieved through Aadhaar.

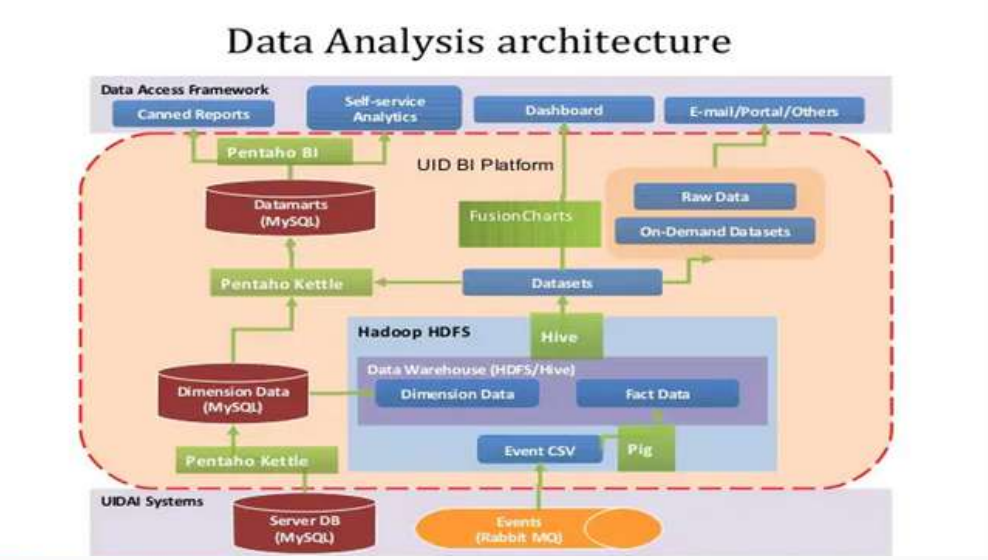
Aadhaar number is for the purpose of identifying individuals for better targeted government subsidies. Since the inception of the UID programme and particularly in the wake of the various e-governance schemes under the National e-Governance Plan, and its successor, the Digital India programme, we see continued attempts to link Aadhaar with various programmes. In an attempt to make these schemes available to the citizens of India almost all citizens enrolled for the same, which resulted in generation of huge amount of data. Handling of this humongous data becomes a challenging task.

Another problem identified was with respect to sharing of databases as the information is confidential. Data sharing usually includes data matching, joint access to repositories of data, file duplication, and data access mechanisms that allow more than one agency or organization to use personal data. Such data sharing results in personal data moving out of traditional data silos and being used in new ways, by different agencies, or for new or different purposes.

Because the data collected for the UIDAI project is sensitive and serves as a gateway to a variety of benefits, backups are required in the event of data corruption or crashes. This brings in the issue of data security and governance.

SOLUTION

The solution to the above problems lies in Big Data Analytics. Earlier, there was no communication between the databases. In the Ginger Platform of the UID project, the service providers have to move their existing databases onto the Ginger platform. The Ginger platform, then organized the available and new data according to the Aadhar number. This integration of database was enabled by Apache Sqoop.Using the Sqoop utility, integration of RDBMS data with HDFS and vice-versa was ensured. After the data is organized, queries can also be executed on the existing data. Apache Hive is being used in the Unique Identity Project for building the UIDAI data warehouse and provides the facility of running queries on the data.



Tools: Apache Sqoop, Apache Hive, Map Reduce

Harshrim Pardal
2028160

COMPANIES USING BIG DATA

"Companies Using Big Data Technology"

AMAZON

Amazon is a renowned online shopping site. They save every bit of information on their clients in order to figure out how they spend their money on a certain product. All of this data is being gathered for use in social media advertising algorithms, which can then be used to grow customer relationships, propose items, improve customer experience and services, and so on.
For example, you may have observed that when you put anything in wishlist or add a product to your basket, it suggests other goods linked to your product or displays items purchased combined with that product.
In this approach, Amazon leverages big data to promote rapid purchases from customers while also improving the overall shopping experience. Big Data is also used to manage product prices in order to attract more customers while increasing net profit.

AMERICAN EXPRESS

The American Express Company analyses and predicts consumer behaviour using big data. In place of typical business intelligence-based hindsight reporting, the organisation applies complex prediction models based on prior transactions and including more than 100 factors. This enables a more precise prediction of possible churn and client loyalty. In fact, American Express claims to be able to anticipate 24 per cent of accounts that would terminate within four months in their Australian market.

BDO

BDO, a national accounting and audit business, uses big data analytics to detect risk and fraud during audits. Previously, discovering the source of a discrepancy would have required multiple interviews and hours of manpower; now, examining internal data first allows for a much narrower field and faster approach. According to BDO Consulting Director Kirstie Tiernan, in one situation, they were able to reduce a list of thousands of vendors to a dozen and then evaluate data individually for irregularities. Fortunately, a specific source was readily located.

CAPITAL ONE

Marketing is one of the most frequent applications for big data, and Capital One is on uppermost of the game, leveraging big data management to guarantee the success of all consumer products.

Capital One identifies the best times to give various offers to clients based on demographics and spending behaviours, enhancing conversion rates from their communications.

Not only does this progress acceptance, but it also makes marketing campaigns significantly more focused and relevant, resulting in better budget allocation.

"Companies Using Big Data Technology"

"Companies Using Big Data Technology"

GENERAL ELECTRIC

Data from sensors on machinery such as gas turbines and jet engines are being used by GE to uncover methods to enhance operating procedures and dependability. The resulting reports are then forwarded to GE's analytics division, which is responsible for developing tools and changes to increase efficiency.
The business estimates that data may increase productivity in the United States by 1.5 per cent, saving enough money to enhance average national incomes by up to 30 per cent over a 20-year period.

MINICLIP

Miniclip, a global developer, publisher, and distributor of digital games, uses big data to monitor and improve user experience.
Because of the nature of the firm and industry, Miniclip prioritises client retention in order to make games more lucrative and, as a result, to promote business development.
Big data reporting, analysis, experimentation, and machine learning data products enable the organisation to quantify the successful portions of their goods and incorporate them in future initiatives while removing or enhancing the problematic components.

"Companies Using Big Data Technology"

NETFLIX

The entertainment streaming service offers a trove of data and analytics that provide insight into the watching habits of millions of foreign customers.
This data is used by Netflix to commission original programming material that appeals to a worldwide audience, as well as to purchase the rights to films and series boxsets that they know would do well with specific populations.
For example, Adam Sandler has been unpopular in the US and UK markets in recent years, but Netflix greenlit four new films with him in 2015, knowing that his prior work had been successful in Latin America.

NEXT BIG SOUND

Next Big Sound (NBS) has found out how to anticipate the next big thing in music using data from Spotify streams, iTunes sales, SoundCloud plays, Facebook likes, Wikipedia page views, YouTube clicks, and Twitter mentions.The company's analytics give insight into social media popularity, the influence of TV appearances, and a plethora of other data points that are vital to the music industry. Thanks to a collaboration between NBS and Spotify, artists may also utilise the data for their own promotion.
Billboard currently produces two charts that are solely based on NBS data, and they have collaborated with firms like Pepsi and American Express to help influence billions of dollars spent by marketers on music-related marketing and sponsorships.

Ishita Bhatnagar
2027541

STARBUCKS

Starbucks is a well-known multinational brand known for producing high-quality coffee. As a result, no single person can manage this volume of data. So, now comes the role of big data.

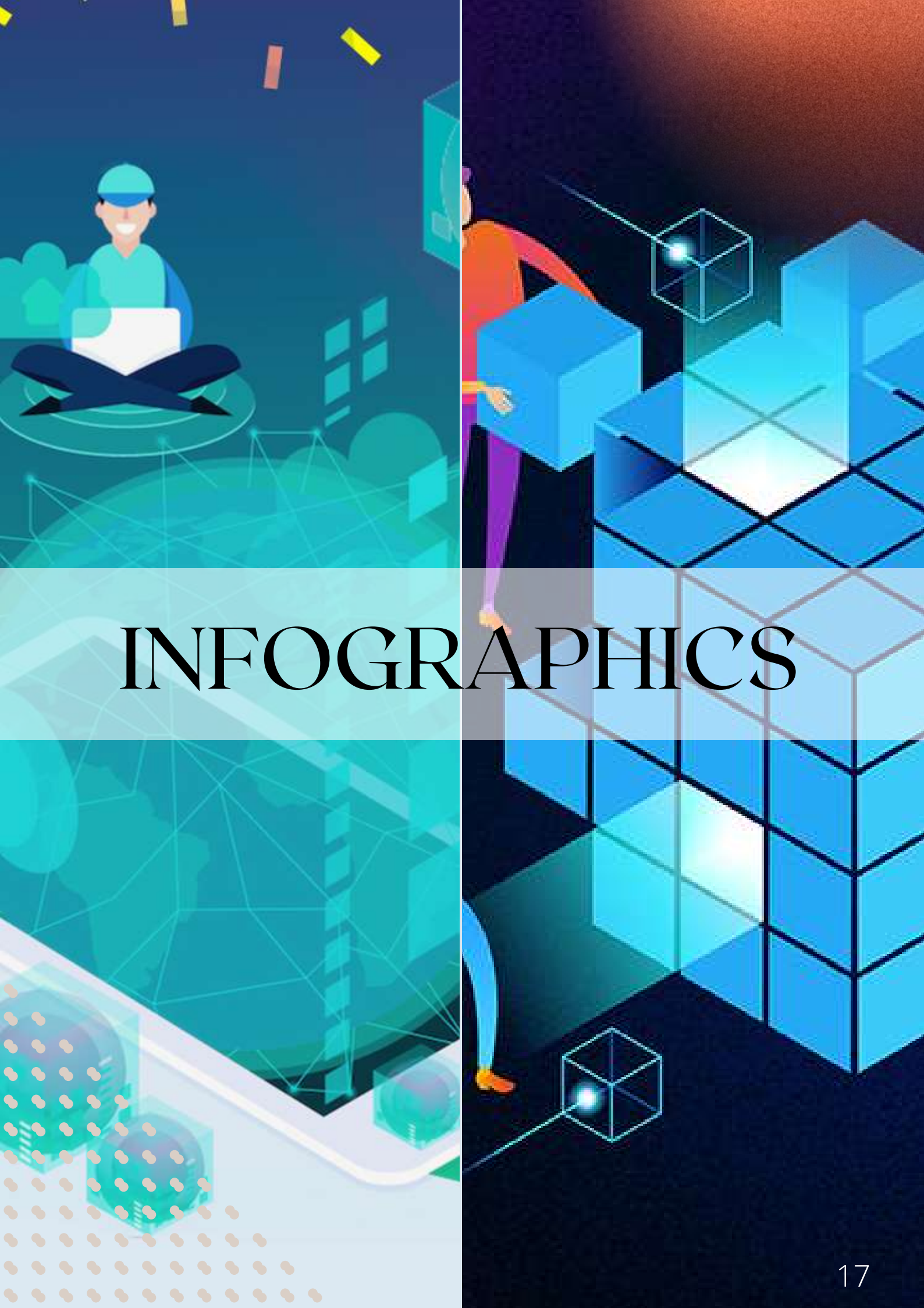
When a consumer uses a mobile app or orders anything, big data captures all of their information. They display the recommended things based on the data obtained.

Big data is also used by the corporation to tailor its menu to the tastes of its customers. They also utilize this data to build more targeted marketing campaigns and promotions, select sites for additional restaurants, and even make future menu changes.

T-MOBILE

Like American Express, the mobile network is using consumer transaction and interaction data to forecast client swings.

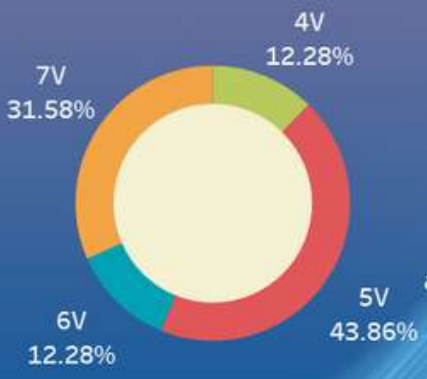
T-Mobile USA says that by combining internal data on billing and customer relations management with data on social media activity, they were able to cut customer defections in half in a single quarter. The data collection technologies have been incorporated into the company's IT infrastructure.



INFOGRAPHICS

How many respondents participated in this survey from different specializations?

How many V's are there in Big Data?



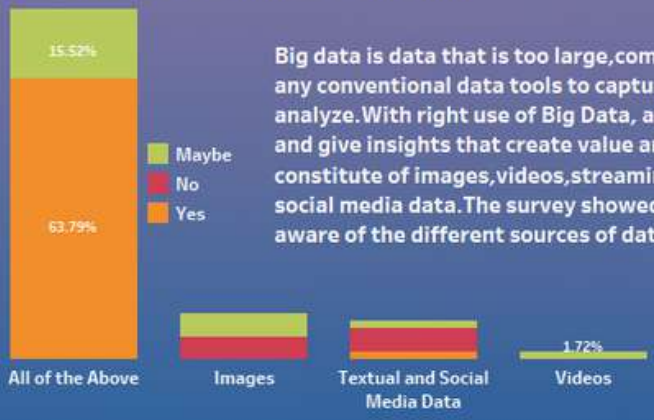
The 7 V's of Big Data:

Volume, Velocity, Variety, Variability, Veracity, Visualization and Value.

From the survey responses it is seen that they are aware of the 7 V's of the Big Data.

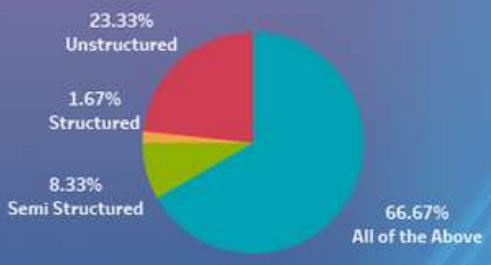
Which company given below is/are implementing Big Data concept?

Have you ever come across the term "Big Data"? What according to you constitute Big Data?



Big data is data that is too large,complex and dynamic for any conventional data tools to capture, manage and analyze.With right use of Big Data, analyst can spot trends and give insights that create value and innovation. Big Data constitute of images,videos,streaming data,textual and social media data.The survey showed that respondents are aware of the different sources of data.

How is data classfied in Big Data?



We generate an unfathomable amount of data every second as the internet age progressess. In order to make insights there is need to first see the structure of the data. There are three types of Big Data:

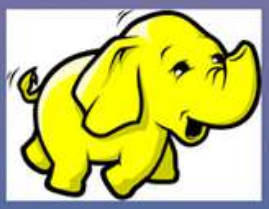
Structured data, Unstructured data and Semi-Structured data.

Survey showed that majority of the respondents are aware of the data types of Big Data.

Why do you think Big Data is important ?

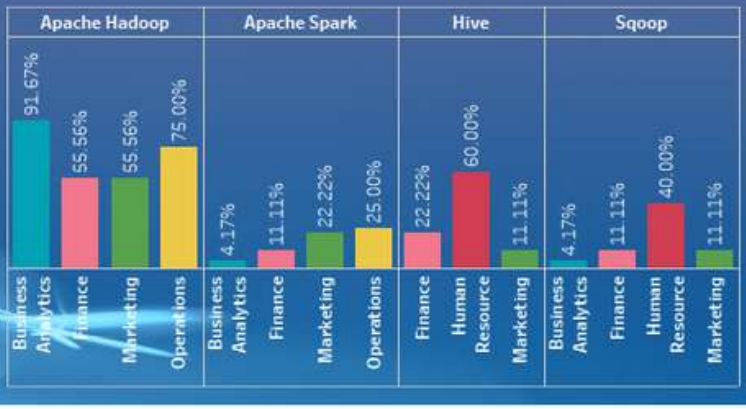
Drive new revenue and growth opportunities	3.33%
Improve operational efficiencies	6.67%
Enable smart decision making.	6.67%
Optimize product development	10.00%
All of the Above	73.33%

Which Big Data tool you would associate with the image given below?

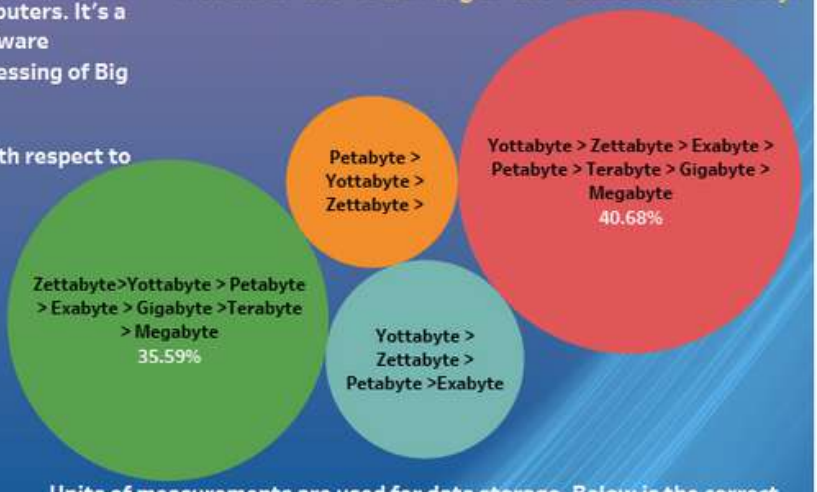


Apache Hadoop is a set of open-source software tools for solving problems involving huge amount of data and computations using a network of many computers. It's a MapReduce programming model-based software framework for distributed storage and processing of Big Data.

Below visualization shows the responses with respect to different specializations .



Which of the following is the correct hierarchy?



Units of measurements are used for data storage. Below is the correct order:

Yottabyte > Zettabyte > Exabyte > Petabyte > Terabyte > Gigabyte > Megabyte

Shweta Seth
2027661

Komal Nagarajan
2027654

BIG DATA ANALYTICS

TOOLS AND TECHNIQUES

Xplenty

Xplenty is a platform to integrate, process, and prepare data for analytics on the cloud. Its intuitive graphic interface will help you with implementing ETL, ELT, or a replication solution.

adverity SMART WITH DATA

Adverity is a flexible end-to-end marketing analytics platform that enables marketers to track marketing performance in a single view and effortlessly uncover new insights in real-time.

Dataddo

Dataddo is a no-coding, cloud-based ETL platform that puts flexibility first – with a wide range of connectors and the ability to choose your own metrics and attributes, Dataddo makes creating stable data pipelines simple and fast.

APACHE hadoop

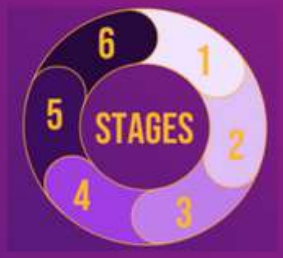
Apache Hadoop is a software framework employed for clustered file system and handling of big data. It processes datasets of big data by means of the MapReduce programming model.

ATLAS.ti

Atlas.ti is all-in-one research software. This tool gives you all-in-one access to the entire range of platforms. You can use it for qualitative data analysis and mixed methods research in academic, market, and user experience research.

PHASES OF BIG DATA ANALYSIS

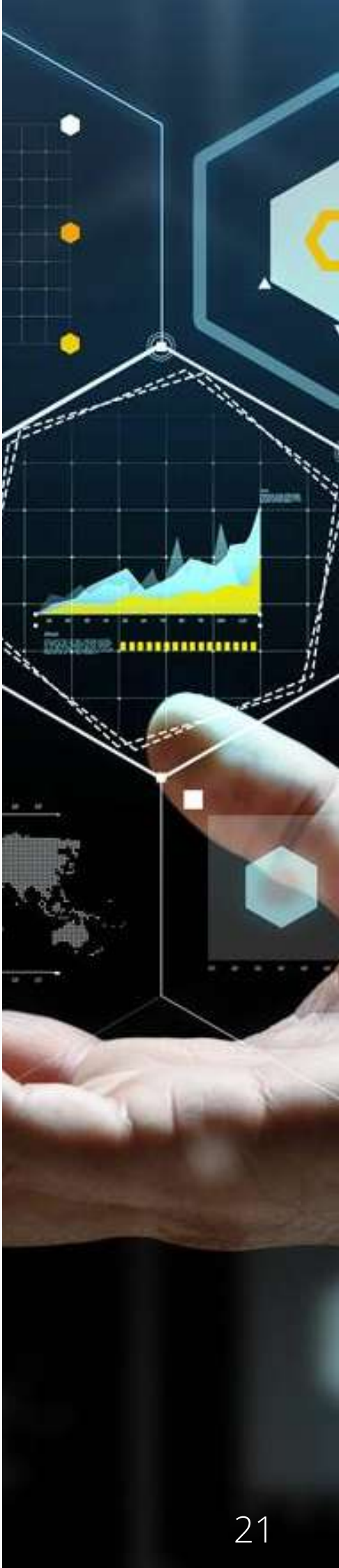
1. BUSINESS CASE EVALUATION
2. IDENTIFICATION & FILTERING
3. DATA EXTRACTION
4. DATA AGGREGATION
5. ANALYSIS AND VISUALIZATION
6. FINAL ANALYSIS RESULT



SAKSHI SH
2127

Sakshi Sharma
2127559

WHITE PAPERS



"Whitepapers on Big Data Technology"

Fraud Detection in the Healthcare Insurance Industry

Companies handle a lot of data daily to thrive. These data originate from a variety of sources and are in a variety of forms. Furthermore, this data comprises some of the most critical aspects of the company's future operations. As a result, businesses must store, handle, and, most importantly, safeguard data. Without data security, a lot of information may be utilized by other firms or, worse, stolen.

Insurance companies face a significant and costly problem in the form of fraudulent claims, which might result in billions of dollars in unwarranted costs each year. To create the basis for false claims, insurance fraudsters frequently exaggerate or manufacture incidents. Traditional insurance fraud detection procedures are time-consuming and complicated. Because of the large number of false claims paid out by insurance companies, premiums have been raised by hundreds of dollars to compensate for the fraudulent payouts, lowering insurance businesses' competitiveness and quality of service. As a result, there is a compelling need to provide quick and effective fraud detection, risk measurement, and safe data management systems that strike the right balance between client personal data preservation, loss prevention savings, and false alarm detection expenditure.

The National Healthcare Anti-Fraud association estimated the total cost of healthcare fraud to be around $80 billion annually. This is probably equal to $200 billion of total spending on healthcare. To prevent insurance fraud, big data solutions are being deployed that use business rule technologies, database searches, anomaly detection and social network analysis.

The Romanian Health Insurance agency uses patient data, data from healthcare providers like doctors, hospitals, and pharmacies, and data from insurance providers. The company uses data mining to classify, cluster and segment data that enables in finding of associations and patterns to detect different types of fraud.

Sai Sandeep Bhoslay
2027032

Pranav KR
2027013

Roshan Kirpalani
2027025

Machine learning algorithms and neural networks can now check every insurance transaction and score its suspicion level.

The Romanian Health Insurance agency developed a scalable, distributed solution based on Hadoop, Apache Spark and TensorFlow. This system automated around 95% of medical claims including ambulatory, dental, hospitalization, and pharmacy claims. It also enabled health insurance agencies to reduce claim's turnaround time and operational costs and thus helped detect greater fraudulent claims within 60 seconds of deployment. High performance computing clusters connected to grid computing infrastructures are the most frequent platforms for running software frameworks that aid big-data processing. In this situation, a cloud computing system uses virtualized storage technologies and delivers dependable services. These platforms can operate as a computer to analyze and interpret the data, and as a web-based visualization tool for the user. The almost free to use Open-Source Hadoop system on economic community hardware is a perfect solution. Hadoop prioritizes performance above ACID compliance. It's also built to handle massive amounts of data.

The challenges with Big Data utilization are information overload and identifying relevant data. The possibilities for false positives and inaccurate interpretation can hurt the whole system. Many firms, such as insurance companies, did not even consider fraud detection as a solution until recently due to the high cost and time it needed to execute the operation. However, the evolving capabilities of Big Data and analytics have resulted in new, cost-effective solutions for identifying unusual transactions and claims that outweigh the traditional costs of fraud identification techniques. With the right technology and architecture, the vast unstructured healthcare data can be sourced, managed, and utilized to improve the quality of service.

"Whitepapers on Big Data Technology"

BIG DATA ANALYTICS IN TRANSPORTATION

INTRODUCTION

The Indian transportation sector could benefit from leveraging the strengths and capabilities of the Internet of Things (IoT) and employing big data analytics.The sensory and communication abilities of the connected physical electronic things also help with IoT. The environment is monitored and reported through these objects, and communicating electronic objects are programmed to act in response to the information received. The intercommunication of physical electronic items generates diverse, large-volume, and high-velocity data (big data).As a result, it becomes vital to investigate the analytics of the created large data for prescription, description, prediction, and actionable insights.

ADVANTAGES OF USING BIG DATA ANALYTICS

- Big data analytics (for example, Hadoop) has numerous benefits for the logistics and transportation industries, including:
- More efficient for real-time big logistics applications involving a large network of sensors and GPS devices.
- Creation of powerful logistics projects with wiser tactics based on collected and processed data
- Creation of effective real-time traffic monitoring applications

PROBLEM STATEMENT

The major goal is to use cutting-edge technology such as Big Data analytics, Hadoop and IOT to increase operational efficiency for logistics and transportation companies. Indeed, these technologies assist company executives in making better business judgments. Here we collect data from automobiles regarding fuel, speed, acceleration, GPS location coordinates, and other data such as date, time, driver's id, and so on using vehicle sensors and GPS devices, and then transfer this information via packets over wireless communication (GPRS) to clustered servers running Hadoop.

Shweta Seth
2027661

RECOMMENDED SOLUTION

The purpose of this project is to create a system that combines internet of things, SaaS cloud architecture, and big data analytics technologies to create an effective real-time monitoring system for customers' cargoes. Mobile phones are used to collect the data, which includes 2D QR codes, GPS positions, and RFID electronic codes. The data is then transferred across wireless networks using the RSA encryption algorithm to preserve the privacy of clients. All of this unstructured data is stored in a big data analytics system that uses HBase as a database.(*X. Lin, X. Zheng,)*

The shipping business is currently quite active, with a huge number of containers to be moved every day. Unique identifying codes are put on each container to monitor delivery, however human reading of these codes has a number of drawbacks, including sluggish speed and a high error rate. Furthermore, unlike vehicle license plate recognition systems, container-code identification systems face greater difficulties due to the low code/container contrast and the wide range of sizes, colors, locations, inter-spaces, and alignment of these codes.Text detection, character extraction, and ultimately text recognition are the three processes in automatic container code recognition. We chose a robust method insensitive to code contrast and other text variable characteristics, such as a texture-based text detection method using Haar wavelet transform for text features extraction and Support Vector Machine (SVM) to classify these features into text and non-text regions, because the text detection step is critical for the other steps. Here we use Hadoop MapReduce to have a parallel execution model as a solution to lower the calculation time for the proposed system to overcome the high computation time of the employed text detection and text recognition methods, which is a huge difficulty for real-time and industry applications like ship transportation.To begin, container code is collected using surveillance cameras or mobile devices and saved to a Hadoop distributed system file (HDFS).

The captured image is then subjected to pre-processing and color image greying procedures. The grey image is then decomposed into 20x20 pixel blocks. To extract the text sections, these blocks are examined and categorized individually on multiple machines using the MapReduce programming technique. The code characters are then separated from the retrieved text sections in a subsequent stage. Then, using the MapReduce programming model, individual characters are subjected to Optical Character Recognition (OCR). Finally, these characters are combined to form container code.

"Whitepapers on Big Data Technology"

BIG DATA FOR WILDFIRE FORECASTING

INTRODUCTION

A wildfire is an unplanned fire that burns in a natural area such as a forest, or grassland. Wildfires are often caused by human activity or a natural phenomenon such as lightning. One of the key reasons for a surge in wildfires is climate change which is again induced due to human activities. The warmer and drier conditions increase tree mortality and fire hazards. Studies have also shown climate change affects the number of wildfire occurrences and increases wildfire intensity and length of the season.

NEED FOR SOLUTION

It can be concluded from the existing studies that there is a huge impact of the wildfires. The wildfires caused due to climate change burn more forest area. They lead to loss of wildlife, forest area, pollution, health impact etc. The existing work focuses on analysing the risk and impact caused by the wildfire. However, there has been no useful step in the direction of preventing such fires. One of the untapped uses of data for wildfires is using it to forecast the wildfire. Billions of lives, money and forest area can be saved if a wildfire can be forecasted and steps can be taken in order to curb it immediately with the help of officials and workers. Thus, there is a need to forecast the wildfire before it brings in huge devastation from an economic, ecological and environmental perspective.

SOLUTION

The dataset consists of data on i) daily weather data, ii) images of forests, iii) Historical burned areas etc. Data wrangling of data in order to bring it to a consistent format will be achieved through Apache Zeppelin. Hadoop is necessary for this solution as the data being used is voluminous and comes from a variety of data sources and Hadoop plays a crucial role in handling big data. The dataset is then loaded into Spark ML and the ConvLSTM model is trained. A ConvLSTM model helps in effectively capturing the spatio-temporal features from the dataset and helps build a model that combines the strengths of both LSTM and CNN.

Harshrim Pardal
2028160

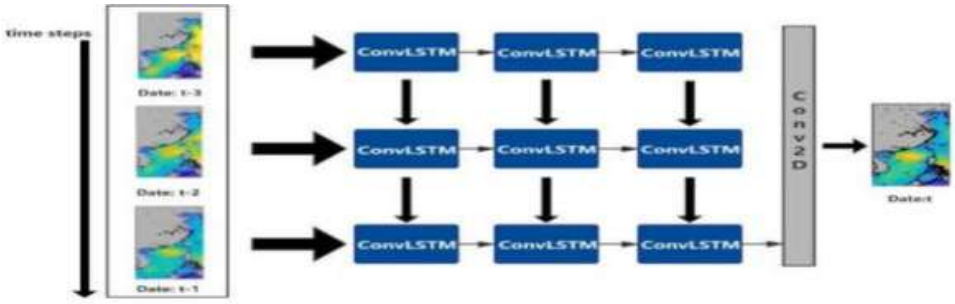
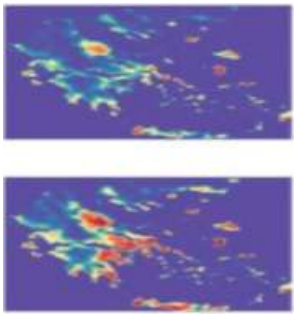


Fig: Proposed Architecture of ConvLSTM based Wildfire Prediction

The ConvLSTM will be used in achieving results such as below for wildfire forecasting.

Fig: Wildfire Forecasted Maps



The sample on the left output shows the forecasted wildfire map of a particular on two different dates. Blue part indicates no danger and extreme danger due to an expected wildfire is showcased through red color. A distributed streaming engine like Kafka or MapR Streams could play an important role in operationalizing the model because the stream ensures that requests and responses can be saved, replicated, and re-playable forever. To compute the wildfire probabilities for the entire landscape, partitioning of the fire simulations will be applied. This will be done by executing several simulations parallely in the Cloud. By applying simulations in the Cloud, the execution time will reduce significantly. Each simulation will be running in a different VM with a fraction of the total targeted number of fires and thus create a wildfire forecast map. Cloud Security will be maintained by implementing least privilege tactics, encrypting the data and monitoring the environment. Strategies such as creation of a privacy task force, establishment of frameworks to assess risk exposure can help data providers take a proactive approach to tackle privacy risk.

CONCLUSION

In this proposed solution, a unique Data Mining approach for forecasting wildfires that incorporates historical data, satellite data, and other sources is proposed. The requisite processing power and storage will be provided through the Cloud. The proposed solution can be expanded in the future by also predicting the scale of the forecasted wildfire. Depending on the forecasted intensity the resources can be sent for its controlling. This will prevent sending either less or more resources and ensures optimal usage and allocation of resources.

Puzzle

Anchal Aneja
2027245

ACROSS

1. Name the unit of data that flows through a Flume Agent.
2. 'Big data' is a term coined by.
3. Apache Kafka is an open-source Big data platform developed by the Apache Software Foundation.
4. In Hadoop ecosystem, Job tracker runs on.

DOWN

1. Name the Big data "V" that relates to the ability to turn your data into relevant business information.
2. What is the name of the Hadoop module that imports and exports data between HDFS and RDBMS?
3. What is the name of the Google-developed programming framework that enables the creation of applications for processing Big data sets in a distributed computing environment?
4. What is the name of an open source framework for storing data and running applications on commodity hardware clusters?

U	B	M	B	S	L	F	G	B	E	V	E	N	T
Y	T	Z	S	U	A	I	R	S	H	A	T	S	K
X	P	R	Q	I	K	L	O	N	D	L	T	M	L
A	V	G	O	N	G	S	M	R	X	U	C	D	T
G	W	J	O	H	N	M	A	S	H	E	Y	P	V
B	L	C	P	T	Z	Q	P	W	X	O	V	A	I
M	Z	T	A	O	H	G	R	I	S	E	K	W	D
L	Z	G	S	R	X	M	E	F	S	A	Y	U	A
V	P	L	I	N	K	E	D	I	N	W	G	V	I
A	G	K	E	F	S	A	U	Y	A	L	V	Y	U
L	H	Z	G	M	J	V	C	T	K	G	H	W	M
N	A	M	E	N	O	D	E	Z	O	I	P	B	S
S	D	L	R	G	V	M	Z	X	T	W	L	F	C
P	O	B	N	M	A	L	A	V	Y	U	A	F	L
K	O	E	S	T	W	P	D	M	N	O	P	S	K
L	P	Z	K	E	F	S	A	Y	L	D	X	O	B

Answers
Across: 1.event 2.johnmashey 3.linkedin 4.namenode
down: 1.value 2.mapreduce 3.sqoop 4.hadoop

Newsletter Team

Shweta Seth
2027661

Komal Nagarajan
2027654

Ishita Bhatnagar
2027541

Vishnu C R
2027115

Shiva kumar Patil
2027564

Devashish Manilbose
2027564

Anchal Aneja
2027245

Scan to access previous Newsletters