Advances in Data Analytics Technologies & Application

Group Project Report

IEDE Spring 2023 by Prof. Kris

Tsinghua University

Team Bios

Name: KELVIN AMOS NICODEMAS

Position: Team leader

Degree: Bachelor

Major: Computer Science & Technology

University: Nanjing University of Information science and technology

Nationality: Tanzania

Name: Seid Seid Aliyu

Degree: Master's

Degree Major: Control Science and Engineering

University: Central South University

Nationality: Ethiopia

Name: Thanapong Ngarnrungruang

Degree: MBA

Degree Major: Masters of Business Administration/ BBA Finance/ Business Analytics

University: Xi'An Jiaotong University

Nationality: Thai

Name: EL AOUAD CHAIMAA

Degree: Bachelor

Degree Major: Artificial intelligence

University: Nanjing University of Information Science and Technology

Nationality: Moroccan

Name: NAIMUL ISLAM MEHEDI

Degree: Bachelor

Degree Major: Artificial Intelligence

University: Nanjing University of Information Science and Technology

Nationality: Bangladesh

Name: ALI MD SHAHIN

Degree: Bachelor

Degree Major: Computer Science & Technology

University: Nanjing University of Information Science and Technology

Nationality: Bangladesh

Name: SALMAN MD

Degree: Bachelor

Degree Major: Computer Science & Technology

University: Nanjing University of Information Science and Technology

Nationality: Bangladesh

Name: ISLAM TAWHIDUL

Degree: Bachelor

Degree Major: Bachelor of Electrical Engineering & Automation

University: Nanjing University of Information Science and Technology

Nationality: Bangladesh

No.	**Fullname**	**Majors/ Expertise**	**IEDE ID**
1	MR. THANAPONG NGARNRUNGRUANG	Masters of Business Administration/ BBA Finance/ Business Analytics	2023008
2	MR. KELVIN AMOS NICODEMAS	Bachelor of Computer Science & Technology	2023009
3	MS.EL AOUAD CHAIMAA	Bachelor of Artificial intelligence	2023010
4	MR. NAIMUL ISLAM MEHEDI	Bachelor of Artificial intelligence	2023011
5	MR. ISLAM TAWHIDUL	Bachelor of Electrical Engineering & Automation	2023012
6	MR. ALI MD SHAHIN	Bachelor of Computer Science and Technology	2023013
7	MR. SEID SEID ALIYU	Masters of Control Science and Engineering	2023014
8	MR. SALMAN MD	Bachelor of Computer Science & Technology	2023015

List of people in to work on different parts of the project

Data collection, analysis and cleaning
Kelvin
Salman(沙满)
Shahin Ali (李新)

Coding part (Visualization)
Chaimaa - 慧英
Mehedi - 梅赫迪
ISLAM TAWHIDUL

Mathematics problem
SEID SEID ALIYU (李想)
Thanapong Ngarnrungruang (陈懿明)

ABSTRACT

Huge amounts of data are being generated continuously as a result of the digitization of most processes, the emergence of various social network platforms, blogs, the deployment of various types of sensors, the adoption of hand-held digital devices, wearable devices, and the explosion in Internet usage. There is no denying that the internet has altered how businesses, governments, and people around the world live their lives. The rate of data generation is currently very high, and the type of data being generated is beyond the capacity of the currently available data storage techniques. This trend is currently in a transformative stage. There is no denying that the emergence and widespread use of the Internet has increased the amount of information that these data carry compared to earlier times.

Therefore in order to analyse this big data we come up with the term data analytics. data analysis is one of the vital elements to process data. Its goals include research identification, transformation, support for decision-making, and research conclusion. The name of data analysis itself varies depending on the study's domain 1, which can be either business, science, or social science. The data analysis can be done in a variety of ways. which provides information about the research problem to support the research.

The goal of the project is to investigate the most recent advancements in data analytics technology and their applications in various sectors. The development of data analytics techniques and technologies, such as big data analytics, machine learning, and artificial intelligence, will be looked at in the research. The project will give a general overview of various data analytics methodologies and techniques, data sources, technology for processing data, tools for data visualisation and analytics, and applications in diverse fields. Also, the research will look at the problems that data analytics technology may have and possible fixes for them. The study will also look at the ethical issues and restrictions associated with employing data analytics tools in various scenarios.

This project's main objective is to present a current overview of the most recent developments in data analytics technologies and their applications. The knowledge and comprehension of data analytics technologies and their effects on businesses and society will be aided by this initiative. The project's focus will be restricted to reports and published research papers from reliable sources. The study won't require gathering or analyzing first-hand data.

Key Words: Data Analytics, Big Data, Data Visualisation, Processing Data, machine learning and artificial intelligence, Data Sources.

Table of Contents

1. Introduction

Data analytics is the collection, transformation, and organization of data to draw conclusions, make predictions, and make informed decisions. Data analytics is often confused with data analytics. Although these terms are related, they are not exactly the same. In fact, data analytics is a subcategory of data analytics that deals specifically with extracting meaning from data. Data analytics generally includes processes beyond analytics, including data science (using data to theorize and predict) and data engineering (creating data systems).

Data analytics is important in many industries because many business leaders use data to make informed decisions. Sneaker manufacturers can review sales data to determine which models should keep and which should be discontinued, or healthcare administrators can view inventory data to determine whether to order. any medical supplies. At Coursera, we may review registration data to determine the types of courses to add to our service. Organizations that use data to drive their business strategies often find that they are more confident, proactive, and financially savvy.

As information technology spreads rapidly, most data is also born in digital form like transactions on the internet today. According to estimates by Lyman and Varian [1], new data stored in digital communication devices accounted for more than 92% in 2002, while the size of this new data is also more than five exabytes. In fact, the problems to analyze data on a large scale that doesn't happen all of a sudden but has been around for several years because it's often much easier to generate data than to find useful things from data. Although today's computer systems are much faster than 1930, large scale data was a limitation to be analyzed by the computers we have today.

To deal with large-scale data analysis problems, several methods are effective [2], such as sampling, data condensation, density-based methods, grid-based methods approach, divide and conquer, incremental learning, and distributed computing, have was presented. Of course, these methods are continuously used to improve the performance of data analysis process operators. The results of these methods are illustrated that with existing efficient methods, we can analyze data on a large scale in a reasonable time. The dimensionality reduction method (e.g. principal component analysis; PCA [3]) is a good example for reducing the volume of input data. speed up the data analysis process. Another shrinking method shrinks data compute clustering of the data being sampled [4], which can also be used to speed things up data analysis computation time.

The size, variety, and rapidly changing nature of this data require a new type of big data analysis, as well as different methods of storage and analysis. Such wholesale quantity data must be analyzed correctly and relevant information must be extracted. Modern businesses place a special emphasis on data and analytics because they can enhance the outcomes of all types of decisions (macro, micro, real-time, cyclical, strategic, tactical, and operational). In addition, D&A can uncover opportunities and questions that business leaders had not even considered, as well as new questions and creative solutions.

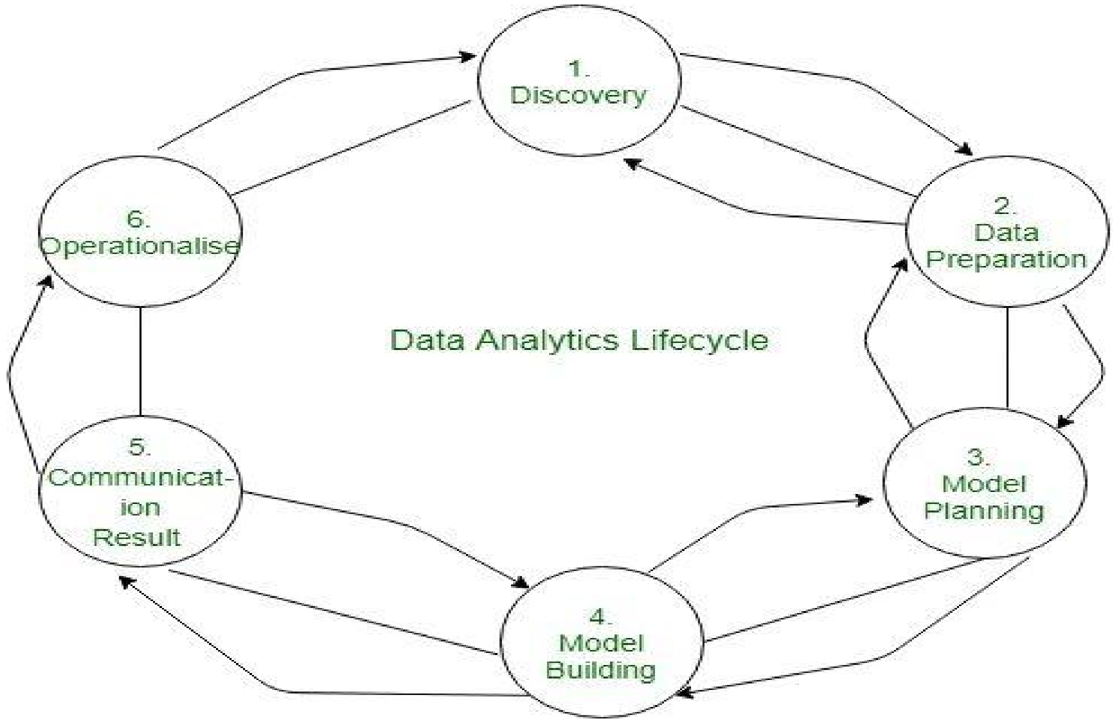


Fig. 1. Data analytics lifecycle

We see the steps for data analytic cycle from figure 1 above, in addition The Data Analytics Lifecycle is designed for big data problems and data science projects. The cycle is repeated to represent an actual project. To meet the unique requirements of big data analytics, a step-by-step approach to organizing activities and tasks related to data collection, processing, analysis, and reuse is required.

Big data and advanced analytics have become the top concern of business leaders around the world for very simple reasons. It will determine the difference between winners and losers in most of our industries in the future. Possibility to increase results in terms of marketing dollars, the possibility of getting additional profit on price and promotion decisions, better ability to achieve additional sales through the supply chain and management decisions - all of these begin to define the gap between you win and loser. So interest rates are very high. The challenge is how to get started and how to actually pursue these ideas and turn them into something more than a hearty dinner speak or speak at major industry conferences, and turn them into things instead Companies act every day.

The data mining, data cleaning, data integration, data transformation and reduction operator can be considered as preprocessing of data analysis [5] trying to extract useful data from raw data (aka key data) and refine them so that they can be used in further data analyses. If the data is a duplicate, incomplete, inconsistent, noisy, or outlier copy, these operators have to clean them. If the data is too complex or large to manipulate, these operators will also try to reduce them. If the raw data contains errors or omissions, the role of the operator must identify them and make them consistent. It can be expected that these operator can influence the analysis result of KDD whether it is positive or negative. Summary, Systematic solutions often include reducing data complexity to speed things up KDD calculation time and improve the accuracy of analysis results.

1.1. Research Background

Big data analytics help individuals make the best decisions for their businesses and institutions. Perfect decision prediction requires accurate and reliable data. Therefore, big data analytics is where decision makers use machine learning to analyze large amounts of data collected from a user's digital footprint and analyze customer preferences, current market trends, contact information for push marketing, etc. The process of extracting valuable information from According to Tseng et al. (2022), Big Data Analytics includes effective approaches to data storage, management, and analysis, leading to key information extraction and intelligent decision making. Big data analytics can facilitate operational decision-making and enable sustainable supply chain systems by implementing various kinds of tools such as machine learning and artificial intelligence. However, the adoption of big data analytics had some potential risks that could impact the sustainability of the supply chain. To pave the way for seamless deployment of big data analytics in enterprises, especially in developing regions, we need to recognize these risks and find ways to mitigate them [6]. Customers' choices and preferences are rapidly changing due to changes in social infrastructure due to rapid technological progress. For the same reason, the industry is facing new challenges and trying to overcome them. fault etc. Whether the vast amounts of data collected by digitally-enabled enterprises can be meaningfully combined with other business resources, transforming vast amounts of data into valuable knowledge that the enterprise can use to survive and improve, whether big data can provide a long-term competitive advantage, and so on. The amount of data extracted from industries using the Internet of Things is growing rapidly, facilitating the practice of artificial intelligence and machine learning. With rich industrial data, industrial design, manufacturing and maintenance results exceed expectations. Big data analytics has become a fundamental tool for enabling intelligent manufacturing systems [9], hospitality and tourism sectors [8].

Big data analytics, machine learning, and data visualization tools referred to as augmentation tools for machine learning are used by Zytek et al. in their qualitative analysis in 2021. To assess the accuracy of high stakes decisions, augmented machine learning predictions are made using these augmentation tools. Making important decisions using big data analytics is frequently given to professionals without any prior experience. This leads to a number of conflicts in the field, including the reduction of complex problems to the output of a single algorithm, vast disparities between human perception and machine learning output, a lack of user confidence in predictions made automatically by machine learning augmentation tools, and finally ethical concerns with big data analytics in high-stakes decision forecasting (i.e., oversimplified algorithms).

Forecasting high stakes decisions entails making a decision that is extremely important, whether it be for personal or professional reasons. High stakes decisions require more time and sophisticated analytical skills than regular decisions do. Due to the possibility of significant financial or emotional loss and the high cost of changing high stakes forecasting, high stakes decisions should not be changed as frequently as they once were. Because most machine learning algorithms are unpredictable, forecasting high-stakes decisions using big data analytics tools presents a number of fundamental challenges. Because the sources of predictions are not obvious to human users, decisions generated by machine learning are difficult for humans to comprehend and interpret [10].

The validity of high stakes decisions is critical and complex in order to maintain the decision's credibility, according to Caines et al.'s(2014)analysis of high stakes decision forecasting at the individual level. In order to choose the best decision for this study, a framework built on the justification of individual decisions provides test scores for various decisions. According to Gati and Kulcsár (2021), the rise of the Internet of Things in the twenty-first century has made career decision-making more difficult. Gati and Kulcsár therefore use three models normative, descriptive, and perspective for three decisions. The normative model shows how individuals should make decisions, the descriptive model shows how individuals actually make decisions, and the perspective model shows how individuals may reasonably make career decisions. Phan et al. (2021) analyze whether deciding bodies rely on arbitrarily chosen or an asymmetrical dispersion of returns and incorporate three fundamental behavioural finance concepts: anticipated utility, optimum expectation concept, and cumulative prospect concept to evaluate decision-forecasting in finance. The results show that participants used algorithms to select random numbers, which ultimately led to non-random choices. Ilyina et al. (2019) used an interactive algorithm-based universal mechanism, adaptable to any company type and organizational context, to support adaptively changing organizational decision making. Nevertheless, this process reduces the complexity of high-stakes decision-making and improves the effectiveness of managerial decision-forecasting.

Data analytics has some uses in corporate decision forecasting, but there isn't much research on big data analytics' utility in high-stakes decision contexts. Personal and professional forecasting with high stakes are frequently disregarded and viewed as routine. Decision forecasting can therefore be costly to individuals or management entities both financially and emotionally. Making decisions that minimize risk and uncertainty will be made possible with the aid of research on using big data analytics for high stakes forecasting. The current proposal provides us with a forum to talk about the most recent advancement in business process mining. The following special topic is covered in the Special Issue proposal on "Big data analytics in high stake decision forecasting" as an example but not a comprehensive list.

- Corporate investment decision making using the big data analytics
- High stakes decision making in banking and insurance domains
- Property risk estimation through machine learning
- Edu-analytics: Big data, educational institute, and high stakes decision forecasting
- HRM analytics, big data, and high stakes decision forecasting
- Energy-tech: big data and high stakes decision forecasting in energy industry
- Big data analytics and environmental sustainability
- Smart business decision using big data analytics
- Marketing Analytics, big data and high stakes decision forecasting
- Smart data management system using the big data management

- Market correlation identification among different market stimuli through big data analytics
- Forecasting customer buying preferences through big data analytics
- High stakes decision forecasting in IT industry

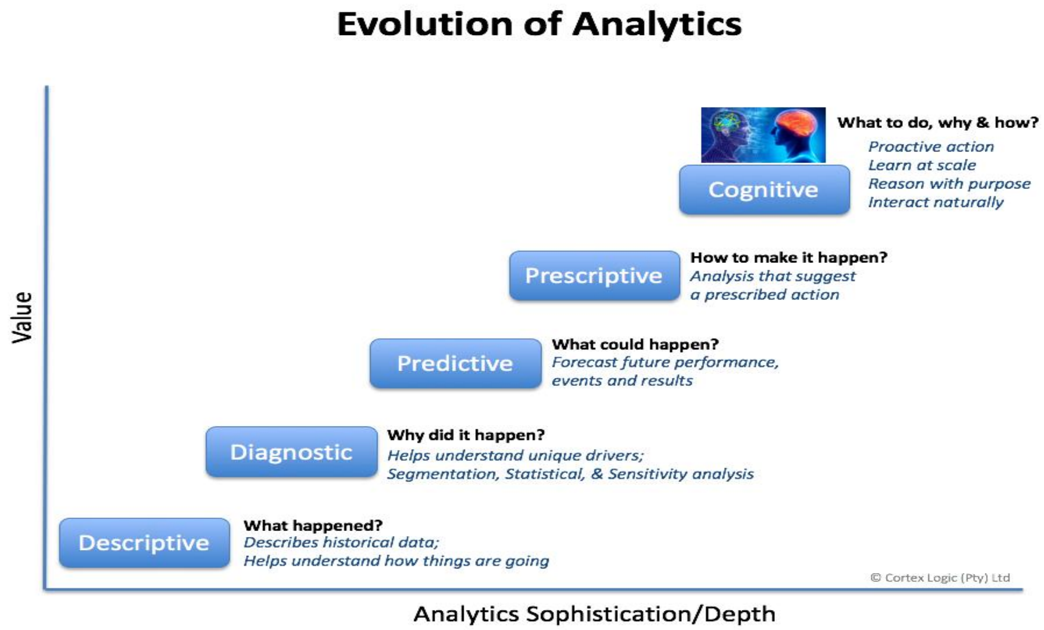


Fig.2. Evolution of data analytics

As we have seen from figure 2 above data analytics go from discriptive to all the way cognitive, and we can see all of the briefly as follows…

Descriptive analytics: Using both recent and old data, descriptive analytics seeks out patterns and connections. Because it only describes trends and relationships without going any further, it is sometimes referred to as the most basic type of data analysis. Your company probably uses descriptive analytics on a daily basis and it is reasonably accessible. Data can be parsed, trends and relationships between variables can be found, and information can be presented visually with the aid of basic statistical software like Microsoft Excel or data visualization tools like Google Charts and Tableau. When communicating change over time, descriptive analytics is particularly helpful. It uses trends as a jumping off point for additional analysis to inform decision-making.

SOME EXAMPLES OF DESCRIPTIVE ANALYTICS

1. Trafic and Engagement Reports
Reporting is a type of descriptive analytics, we already use descriptive analytics if our organization monitors engagement through social media analytics or website traffic. These reports are made by comparing current metrics to historical metrics and visualizing trends using raw data that is generated when users interact with your website, advertisements, or social media content. You might be in charge of reporting on which media outlets bring in the most visitors to the product page on your company's website, for instance. You can count the number of visitors from each

source by examining the page's traffic statistics using descriptive analytics. You might choose to go a step further and contrast current traffic source data with earlier traffic source data.

Fig. 3. Basic principle of discrictive analytics

This will allow us to inform your team of any changes, such as highlighting a 20 percent year-over-year increase in traffic from paid advertisements. The three additional analytics types can then be used to ascertain whether traffic from each source has been increasing or decreasing over time, whether trends are expected to persist, and what your team should do next.

2. Financial Statement Analysis

Financial statement analysis is one more descriptive analytics example that you may be familiar with. Financial statements are periodic reports that include specific financial data about a company and provide an overall picture of its financial health. The balance sheet, income statement, cash flow statement, and statement of shareholders' equity are just a few examples of the various financial statements. Each targets a particular group of people and communicates various financial information. There are three main approaches to financial statement analysis: vertical, horizontal, and ratio. Reading a statement from top to bottom and contrasting each item with those above and below it is known as vertical analysis.

This aids in establishing the connections between variables. If each line item represents a percentage of the total, for example, comparing them can reveal which items account for greater and lesser shares of the total. Reading a statement from left to right and comparing each item to itself from a previous period is known as horizontal analysis. Analyses of this kind track changes over time. Finally, ratio analysis compares different report sections based on how they relate to one another as a whole. This determines whether your business is over- or under performing by directly comparing data from different time periods and your company's ratios to those of the sector. Each of these techniques for analyzing financial statements is an illustration of descriptive analytics because it uses both recent and past data to reveal trends and relationships between various variables.

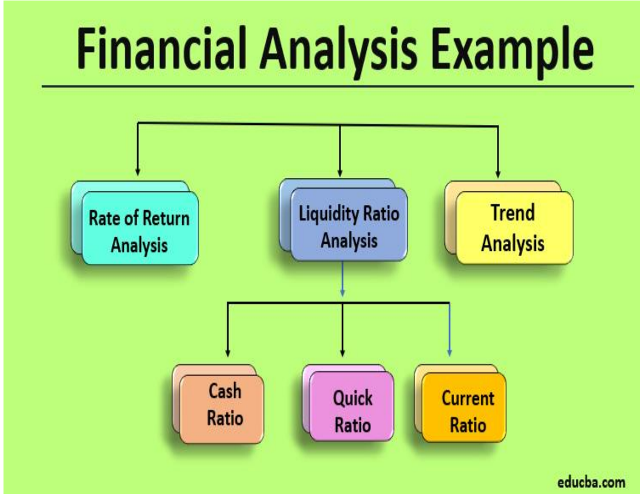


Fig. 4. Financial statement analysis data analytics

3. Demands Trend

Additionally, descriptive analytics can be used to spot patterns in consumer preferences and behavior and predict demand for particular goods or services. A great use case for descriptive analytics is trend identification by streaming service Netflix. The Netflix team, which has a history of being highly data-driven, collects information on users' platform usage. They use this data to analyze what TV shows and movies are popular right now, and they display those titles in a section on the platform's home screen. This information not only enables Netflix subscribers to see what's popular and, consequently, what they might enjoy watching, but it also enables the Netflix team to understand which media genres, themes, and actors are particularly favored at a given time. This may influence choices regarding the creation of new original content, agreements with current production companies, marketing, and retargeting campaigns.

4. Aggregated Survey Results

Market research can also benefit from descriptive analytics. Descriptive analytics can be used to find connections between variables and trends when extracting insights from survey and focus group data. You might, for instance, carry out a survey and find that, as respondents' ages rise, so does their propensity to buy your product. Descriptive analytics can determine whether this age-purchase correlation has always existed or whether it was something that only happened this year if you've conducted this survey repeatedly over a number of years. This kind of information can open the door for diagnostic analytics, which can explain why certain variables are correlated. Using predictive and prescriptive analytics, you can then use the trends to plan future product improvements or marketing campaigns.

5. Progress to Goals

Descriptive analytics can also be used to monitor goal progress. Your team can determine whether efforts are on track or if changes need to be made by reporting on progress toward key performance indicators (KPIs). If your company wants to reach

500,000 unique page views per month, for instance, you can use traffic data to show your progress. You have 200,000 unique page views, so you're probably halfway through the month. You should be halfway toward your objective at that point—at 250,000 unique page views—so this would be under performing. With the help of this descriptive analysis, your team can determine what needs to be changed in order to increase traffic and get back on track to meet your KPI.

Diagnostic analytics: The process of using data to identify the reasons behind trends and correlations between different variables is known as diagnostic analytics. we could think of it as the logical progression from using descriptive analytics to find trends. Manual analysis, algorithmic analysis, and statistical software (like Microsoft Excel) are all options for performing diagnostic analysis. Before beginning diagnostic analytics, it is important to understand a number of concepts, including diagnostic regression analysis, hypothesis testing, and the distinction between correlation and causation.

Diagnostic analytics is the secret to helping businesses that gather customer data comprehend why customers behave the way they do. These revelations can be applied to enhance brand messaging, user experience (UX), and product-audience fit. If we see the HelloFresh example and think about the importance of customer retention to the subscription-based business. HelloFresh uses diagnostic analytics to ascertain the reasons why departing customers decide to cancel subscriptions because keeping existing customers is more cost-effective than acquiring new ones.

Predictive analytics: A subset of advanced analytics called predictive analytics uses historical data along with statistical modeling, data mining, and machine learning to forecast future outcomes. Utilizing patterns in this data, businesses use predictive analytics to spot risks and opportunities. Big data and data science are frequently linked with predictive analytics. Companies are currently flooded with data, which can range from log files to images and video and is stored in various data repositories across an organization. Data scientists use deep learning and machine learning algorithms to find patterns and forecast future events in order to gain insights from this data. Neural networks, decision trees, and logistic and linear regression models are a few of these statistical methods. Some of these modeling methods build on earlier predictive learnings to derive new predictive understandings.

Predictive analytics can be deployed in across various industries for different business problems. Like Banking, Healthcare, Human resources (HR), Marketing and sales and Supply chain.

Prescriptive analytics: Prescriptive analytics is a process that examines data and immediately offers suggestions on how to improve business procedures to accommodate various anticipated outcomes. In essence, prescriptive analytics uses the "what we know" (data) to predict potential outcomes and offers the best course of action based on well-informed simulations.

The logical next step after descriptive and predictive analytics is prescriptive analytics. It takes data analytics one step further by taking the guesswork out of it. Additionally, it saves data scientists and marketers time by allowing them to focus on creating highly customized and advantageous user experiences for their audiences rather than trying to figure out what their data means and how to connect the dots.

Cognitive analytics: In order for the business user or data analyst to gain insights from advanced analytics, cognitive data analysis (CDA) automates and adds cognitive processes to data analysis. In the era of big data, where the data is so complex and includes both structured and unstructured data, CDA is especially crucial because it is impossible to manually check all potential combinations. CDA does not just take over the process because it is a cognitive computing system. Instead, CDA engages with the user and gains knowledge from the engagement.

By making inferences from past data and patterns, drawing conclusions based on prior knowledge bases, and then adding these conclusions back into the knowledge base for subsequent inferences, cognitive analytics seeks to mimic the workings of the human brain. Semantics, artificial intelligence algorithms, deep learning, and machine learning are just a few of the intelligent technologies that cognitive analytics brings together. By using these methods, a cognitive application can gradually learn from its interactions with data and people to become more intelligent and useful.

1.2. Development and Research Status

On this research we can answer some questions like What is the role of data and analytics in business? And also we will describe briefly about the application, opportunities and challenges of data analytics. Modern businesses place a special emphasis on data and analytics because they can enhance the outcomes of all types of decisions (macro, micro, real-time, cyclical, strategic, tactical, and operational). In addition, D&A can uncover opportunities and questions that business leaders had not even considered, as well as new questions and creative solutions. Progressive businesses use data in a variety of ways and frequently rely on data that isn't under their direct control to help them make better business decisions. As it enables quicker, more accurate, and more relevant decisions in challenging and quickly evolving business contexts, data and analytics also serve as a catalyst for digital strategy and transformation.

Data-driven decision making is the process of using data to determine how to make decisions that are better. This suggests the use of a decision model, which can incorporate prescriptive analytical methods that produce results that can specify the course of action to be taken. Predictive, diagnostic, and descriptive analytic models are additional options. Notably, decisions influence action but can also dictate when to refrain from acting. By developing a vision of a data-driven enterprise, quantifying and communicating business outcomes, and encouraging data-fueled business changes, forward-thinking organizations are integrating data and analytics into business strategy and digital transformation.

Organizations are increasingly using advanced analytics to solve business problems, but how and whether to use prediction, forecasting, or simulation for the predictive analysis component depends on the problem's nature and complexity. Decision-making is particularly difficult when scaling a digital business and calls for a combination of data science and more sophisticated methods. Organizations can react quickly to shifting requirements and constraints thanks to the combination of predictive and prescriptive capabilities.

Fig. 5. data analysis process

The following are examples of combining the predictive capabilities of forecasting and simulation with prescriptive capabilities:

- Forecasting the risk of infection during surgery and using set guidelines to motivate risk-reduction measures.

- In order to proactively respond to shifting demand throughout the supply chain, forecasting incoming orders for products is combined with optimization; however, historical data that may be insufficient or "dirty" should not be used.

- To quickly evaluate various scenarios and choose the best response strategy for each, optimization and simulation of the segmentation of customers into micro segments based on risk are used.

Additionally, data and analytics are applied in various ways for various decisions. Executive leaders must understand when and why to combine the best aspects of human decision-making with the power of data, analytics, and AI in order to make more effective business decisions. It's important for each organization to define what data and analytics means for them and what initiatives (projects) and budgets are necessary to capture the opportunities.

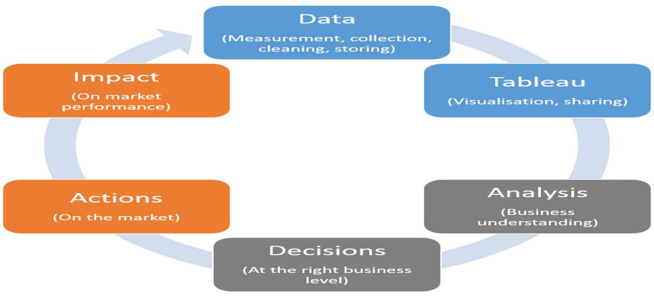


Fig. 6. data analytics planning process

The key steps in data and analytics strategic planning are to:

- start with the mission and goals of the organization
- determine the strategic impact of data and analytics on those goals
- prioritize action steps to realize business goals using data and analytics objectives.
- build a data and analytics strategic roadmap
- implement that roadmap (i.e., projects, programs and products) with a consistent and modern operating model.
- communicate data and analytics strategy and its impact and results to win support for execution.

The enterprise operating model for data and analytics must also work to close gaps in the organizational delivery approaches, architectures, and data ecosystem required to carry out the D&A strategy. The ability to execute actions on the market (by market I mean the environment of the organization in which it operates) and the corporate analytical capabilities, decision-making speed, and ability are all critical to success in the end. The cycle then repeats itself as all of this is reflected in the (performance) data gathered. If the corporate performance has not improved, there may be an issue at various stages:

- ✓ having non-insightful information sources (the data collected or the visualisations) or the lack of deriving business insights.
- ✓ lack of timely business decisions
- ✓ inability to transform the business decisions to actions (or being late with the action).
- ✓ inability to execute the action in the business environment.

1.2.1. Problem of The Study

In this research we will briefly describe the application of data analytics and some of the challenges we have faced in data analytics. Analytics of data is crucial for risk managers.They help employees predict losses and track performance while also enhancing decision-making, increasing accountability, and improving the financial health of the company. Not persuaded? See Why All Risk Managers Should Use Data Analytics and 6 Reasons Data is Key for Risk Management for more information on this topic. But it's easier said than done to get these advantages. Risk managers may encounter a number of difficulties when gathering and applying analytics.

Some of the challenges are…

- ✧ The amount of data being collected
- ✧ Collecting meaningful and real-time data
- ✧ Visual representation of data
- ✧ Data from multiple sources
- ✧ Inaccessible data
- ✧ Poor quality data
- ✧ Pressure from the top
- ✧ Lack of support
- ✧ Confusion or anxiety
- ✧ Budget
- ✧ Shortage of skills
- ✧ Scaling data analysis

The advantages of data analysis make the time and effort required to overcome these obstacles well worthwhile. Consider investing in a data analytics system to improve our business today.

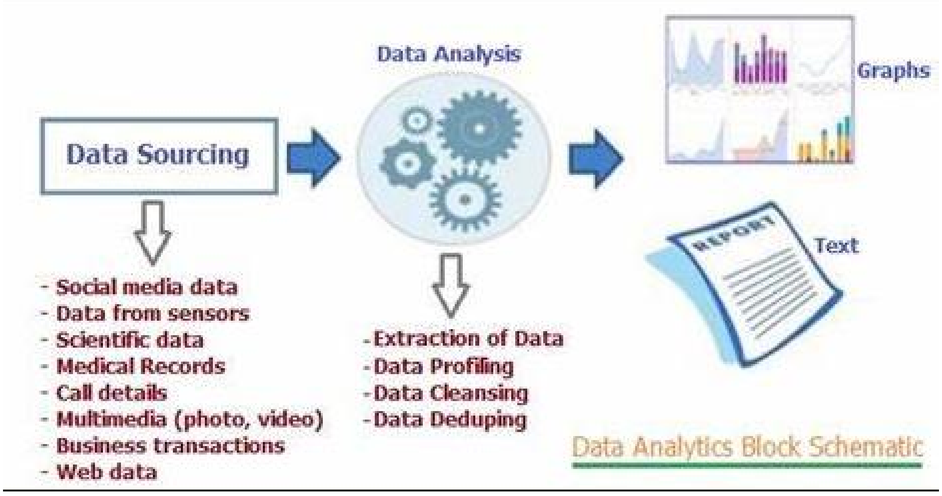


Fig. 7. data analytic problem schematic

Because risk managers and other employees frequently experience data overload as a result of today's data-driven organizations and the advent of big data. Every incident and interaction that occurs within an organization every day may be reported, leaving analysts with thousands of interconnected data sets. A data system that automatically gathers and arranges data is required. Manually carrying out this process would take far too long in the current environment. Employees will be able to use the time normally spent processing data to take action thanks to an automated system.

Data analytics technology has become increasingly important in recent years due to the growing amount of data generated by individuals and organizations. Technology helps individuals and businesses make informed decisions by analyzing large data sets to identify patterns and trends. The purpose of this research is to explore the advances in data analytics technologies and their application. This paper will provide an overview of data analytics technologies and their evolution, data analytics categories and use cases, data sources and tools for processing data, technologies and problems associated with big data analytics, artificial intelligence and machine learning in data analytics, data visualization and analytics tools, the use of data analytics in many businesses, data analytics technologies' challenges, and future directions.

1.3. The Study's Objective and Importance

Objectives:

- To examine the various types of data analytics and how they can be applied in different industries and contexts.

  A wide range of diverse methods and technologies are included in the field of data analytics, which is used to draw conclusions and information from data. Understanding the various forms of analytics and how they can be used in various contexts and businesses is the primary goal of studying data analytics.

  In descriptive analytics, data is summarized and visualized to reveal historical patterns and trends. Analyzing data for diagnostic purposes includes figuring out why specific occurrences or trends happened. In predictive analytics, future events or trends are predicted using data. Data is used in prescriptive analytics to recommend actions or improve decision-making.

  Several sectors and situations can benefit from the usage of these various analytics kinds. Data analytics may be utilized, for instance, in the healthcare sector to spot patterns in patient outcomes and create individualized treatment regimens. Data analytics may be used in the financial sector to find patterns in financial data and make investment decisions. Data analytics may be used in the marketing sector to target certain consumers and improve advertising efforts should be aware of how crucial data analytics is for accelerating digital transformation and making informed decisions.
- Understanding the role that data analytics plays in accelerating digital transformation and informing data-driven decisions is the second goal of studying data analytics.

  Organizations are gathering more data than ever before in the digital age, and data analytics is essential for interpreting this data and generating insights. Organizations may use data analytics to find patterns and trends in their data that will assist them make strategic and operational decisions. Organizations can, for instance, determine the most popular goods and services by studying consumer data and then modify their offerings appropriately. Organizations can find areas of inefficiency and streamline their processes by evaluating financial data.

  Data analytics is essential for accelerating digital transformation, which is the process of transforming conventional business structures and procedures through technology. Organizations may simplify operations, enhance goods and services, and gain a competitive edge in their sector by implementing new technology and data analytics tools.
- To identify the benefits of learning data analytics and gaining certification in this field. The third objective of studying data analytics is to identify the benefits of

learning data analytics and gaining certification in this field. There is a strong need for people with knowledge and experience in this field due to the expanding significance of data analytics across several businesses.

Studying data analytics may provide people an useful set of skills that they can use in a range of situations and sectors. Having a competitive edge in the job market is another benefit, since many companies are trying to recruit people with data analytics expertise.Gaining certification in data analytics can also provide additional. Data analytics tools and methodologies may be learned in an organized manner through certification programs. Also, they can show prospective employers that a person has a high degree of experience in this field.

- To explore the best practices for implementing data and analytics programs. Investigating the top methods for putting data and analytics programs into action is the fourth goal of researching data analytics. There are many aspects to take into account when planning and executing data and analytics initiatives, and it may be a challenging process. Understanding the organization's goals and objectives is a great practice for establishing data and analytics projects. This may make sure that the program is created to fulfill specific business goals and is in line with the organization's broader strategy.

  Making ensuring the program is backed by the appropriate infrastructure and technology is another great practice. This may entail spending money on technology and tools for data analytics as well as establishing the necessary data infrastructure to support these solutions. The best practices for developing data and analytics initiatives also include training and education. For its staff to be able to use data analytics tools and methodologies successfully, organizations need invest in their education and training. This may entail offering well-organized training courses as well as chances for real-world training and experience.

  Another crucial factor to take into account when adopting data and analytics initiatives is data governance. Establishing standards and processes for managing data and making sure it is accurate, consistent, and secure requires data governance. By doing this, it is possible to make sure that the analytics data is reliable and of high quality.

  Successful data and analytics projects also require collaboration and communication. In order to understand the requirements and objectives of other teams within the company and to make sure that analytics insights are successfully conveyed and used to guide decision-making, data analytics teams should collaborate closely with these teams.

Overall, careful planning and taking into account a number of elements are needed to create successful data and analytics projects. Organizations can ensure that their programs are created to address specific business needs, are supported by the appropriate technology and infrastructure, and are staffed with people who have the knowledge and experience required to use data analytics tools and techniques by adhering to best practices.

Importance:

- Data analytics is critical to extracting insights and knowledge from large and varied data sets. It is impossible to exaggerate the value of data analytics. In the current digital era, businesses are gathering enormous volumes of data from several sources, such as social media, financial transactions, and consumer interactions. For making wise judgments and deriving insights and information from this data, data analytics are essential. Organizations run the danger of overlooking significant patterns and trends in their data that might guide company strategy and decision-making if they don't use data analytics. Organizations may increase their understanding of their customers, products, and operations and make better judgments about how to improve and optimize these areas by employing data analytics tools and methodologies.

- Data analytics plays a significant role in various industries, including healthcare, finance, marketing, and more. A range of sectors employ data analytics to enhance company outcomes and decision-making. Data analytics are utilized in the healthcare sector to spot patterns in patient outcomes, create individualized treatment programs, and enhance hospital operations. Data analytics are used in the financial sector to find trends in financial data, choose investments, and control risk. Data analytics are utilized in the marketing sector to target certain consumers and improve advertising efforts. Data analytics will become increasingly important as businesses in all sectors become more data-driven. Data analytics is anticipated to inform company strategy and decision-making considerably more in the upcoming years.

- Learning data analytics can improve decision-making, identify opportunities and risks, and provide a competitive edge in the job market. Studying data analytics may provide people useful skills and experience that they can use in a range of situations and sectors. People may get a deeper grasp of how to extract knowledge and insights from data by learning data analytics. They can then utilize this knowledge to guide decision-making and spot opportunities and hazards. Learning data analytics might also provide you an advantage in the employment market. There is a significant need for people with knowledge and experience in this field because of the expanding significance of data analytics across all businesses. Individuals may show prospective employers that they have a high degree of experience in this field and are able to use data analytics tools and methodologies to produce business outcomes by earning a certification in data analytics.

- By implementing data and analytics initiatives, businesses may boost their competitiveness and accelerate the digital transformation. Employing data and analytics programs can provide businesses a competitive edge in their sector. Organizations may streamline operations, enhance goods and services, and get insights into market trends and consumer behavior by employing data analytics to guide decision-making. Moreover, putting in place data and analytics programs may assist businesses in driving digital transformation, or the use of technology to change conventional business structures and procedures. The importance of digital transformation is rising as more sectors of the economy and companies adopt a digital-first strategy.

By offering insights into how business processes may be streamlined and enhanced via the use of technology, data analytics can play a crucial role in accelerating digital transformation. By using data analytics to identify areas for improvement and optimization, organizations can develop more efficient and effective business processes that are better aligned with their strategic goals. For firms to be competitive in their sector, digital transformation is also crucial. Businesses that choose not to adopt digital technology and procedures run the danger of falling behind and losing market share. Organizations may remain ahead of the curve and set themselves up for long-term success by embracing digital transformation and putting data and analytics initiatives in place.

Generally, firms across all industries must prioritize studying data analytics. By applying data analytics tools and methodologies to extract insights and information from data, companies may enhance decision-making, uncover opportunities and dangers, and achieve a competitive advantage in their sector. Organizations may also accelerate digital transformation, streamline business processes, and position themselves for long-term success in a quickly evolving digital world by deploying data and analytics initiatives.

In conclusion, data analytics is a vital and quickly expanding discipline that is changing how businesses function and make choices. Organizations must make investments in employee education and training, establish data governance policies and procedures, and foster collaboration and communication between data analytics teams and other teams within the company given the growing significance of data analytics across all industries. Also, gaining knowledge of data analytics may help people develop vital skills and knowledge that will provide them an advantage in the job market and increase their ability to make decisions. Adopting data and analytics programs may provide businesses a competitive edge and accelerate digital transformation, both of which are crucial in today's fast shifting corporate environment. The value of data analytics will only increase as we move forward. Businesses that make investments in data analytics now will be well-positioned to prosper in a future where data is used more and more for business results and decision-making.

Understanding what the data contains and does not contain depends on how the data are organized and thought about. There are many different approaches to data analysis, and it is infamously simple to manipulate data during the analysis phase in order to push particular conclusions or agendas.

The outline of this research is organized as follows. **Chapter 1** briefly describes of the **introduction**, chapter 2 is **research review**, chapter 3 describes the **methodology** of the research, chapter 4 **Results and Insights**, chapter 5 **Discussion and Implications**, and finally chapter 6 describes about the **conclusion** of the research.

2. Research Review

The research will review various data analytics technologies and their evolution, including data analytics categories and use cases, data sources, and tools for processing data. It will also examine the technologies and problems associated with big data analytics, artificial intelligence, and machine learning in data analytics, data visualization and analytics tools, and the use of data analytics in many businesses. Finally, the research will discuss the challenges and future directions of data analytics technologies.

2.1. An Overview of Data Analytics Technologies and Their Evolution

A Google paper on the Google File System, published in October 2003, and Jeffrey Dean and Sanjay Ghemawat's MapReduce paper in 2004, kicked off the era of big data technologies. Shortly thereafter, Doug Cutting and Mike Cafarella, then working with Yahoo! on a search engine calledApache Nutch(based on Apache Lucene indexing), created the new Hadoop subproject for running a large-scale computation on large clusters of commodity hardware in January 2006. Since these early efforts, the big data technology landscape has been enriched with numerous innovations and evolved in leaps and bounds. In part one of this blog series, we'll look at the evolution of the data and analytics space across their core aspects.

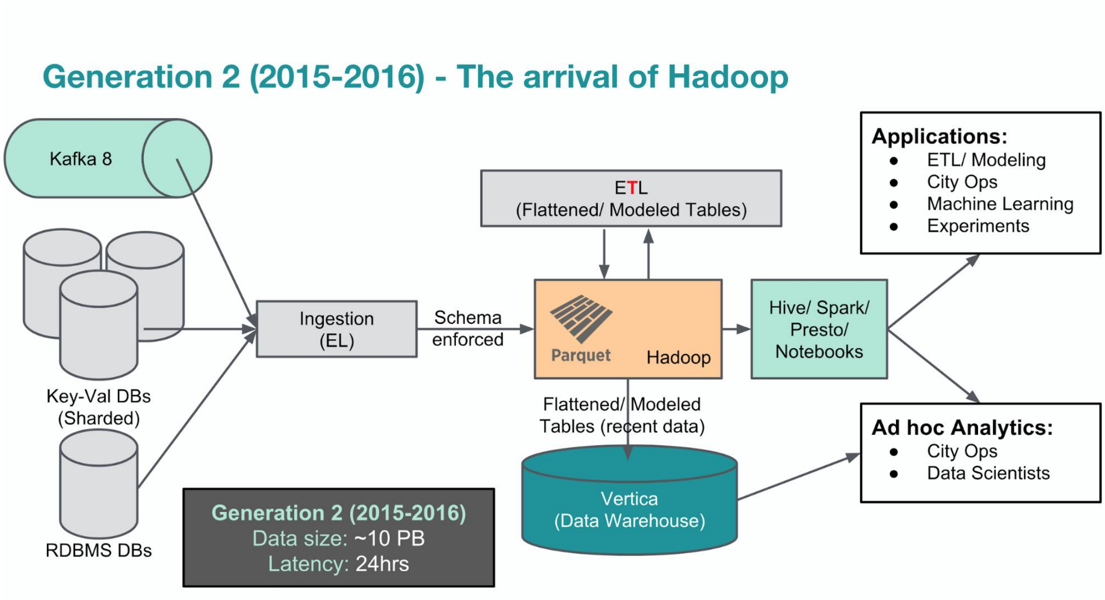
Data Platforms Evolution:

OSS based → Packaged Distributions → Cloud-Native Stack → Cloud Agnostic stack → Composable Data Analytics

Soon after Hadoop released an Apache open source project, it spawned several frameworks on top of Hadoop to perform different types of processing. Apache Pig, Apache Hive, Apache HBase, Apache Giraph, and Apache Mahout were a few of the diverse frameworks that allowed different ways to process data stored in a Hadoop cluster.

In addition, there were parallel stacks that replaced one or more frameworks with Kafka, Cassandra, or Cascading. The initial deployments required teams to build and deploy softwareon commodity hardware based on open-source Hadoop ecosystem components.

After Hadoop's project came the commercial distribution of Cloudera, Hortonworks, MapR, Hadapt, DataStax, and Pivotal Greenplum, which packaged the required software in a user-friendly fashion and provided premium support. Then, Amazon EMR released the first cloud-based Hadoop distribution. Now there are cloud-specific Hadoop and Spark-based distributions like Azure Synapse Analytics GCP DataProc that come pre-installed with the required software and computing power. From there, cloud-agnostic stacks such as Snowflake and Data Bricks evolved to work efficiently across different clouds. These platforms are adding innovative features which cater to key performance and cost metrics. As a result, these technologies are getting quite popular, with many enterprises now moving towards such cloud-agnostic stacks. Enterprises are increasingly looking at a multi-cloud strategy to avoid lock-ins by a particular cloud and use the best technology for various purposes. The trend for the future is to move towards composabledata analytics, where companies willbuild data platformsusing components from two to three different technologies and cloud



providers.

Fig. 8. The first generation of Uber's Big Data platform allowed us to aggregate all of Uber's data in one place and provide standard SQL interface for users to access data.

Data Architecture Evolution:

Data Warehouse → Data Lakes / LakeHouse → Data Mesh / Data Fabric

For decades, data warehouses such as Teradata and Oracle have been used as central repositories for storing integrated data from one or more disparate sources. Thesedata warehousesstore current and historical data in one place that can create analytical reports for workers throughout the enterprise. With the advent of big data frameworks like Hadoop, the concept of a data lake became incredibly popular. Typically, a data lake is a singular data storage and includes raw copies of source system data, sensor data, social data, and transformed data for reporting, visualization, advanced analytics, and machine learning tasks. Data mesh is an organizational and architectural paradigm for managing big data that began to gain popularity in 2019. It is a process and architectural approach that delegates responsibility for specific data sets to business members who have the subject matter expertise to use the data properly. With the data mesh architecture, data domains become prominent with a base data platform that individual domain teams can use with their own data pipelines.

In this situation, data resides within the foundational data platform storage layer (data lake or data warehouse) in its original form. Individual teams will choose how to process this data and then serve the datasets they own in a domain-specific manner. Data fabric is a design concept defined by Gartner that serves as an integrated layer (fabric) of data and connecting processes. A data fabric utilizes continuous analytics over existing, discoverable, and inference metadata assets to support the design, deployment, and utilization of integrated and reusable data across all environments, including hybrid and multi-cloud platforms. Data mesh and data fabric are concepts that provide a unified view of the data distributed across the underlying data lakes and data warehouses. They, however, differ in how users access them. While data mesh is about people and processes, a data fabric is an architectural approach to tackle the complexity of the underlying data. Experts believe these concepts can be used simultaneously and will work on top of the existing data lakes and warehouses.

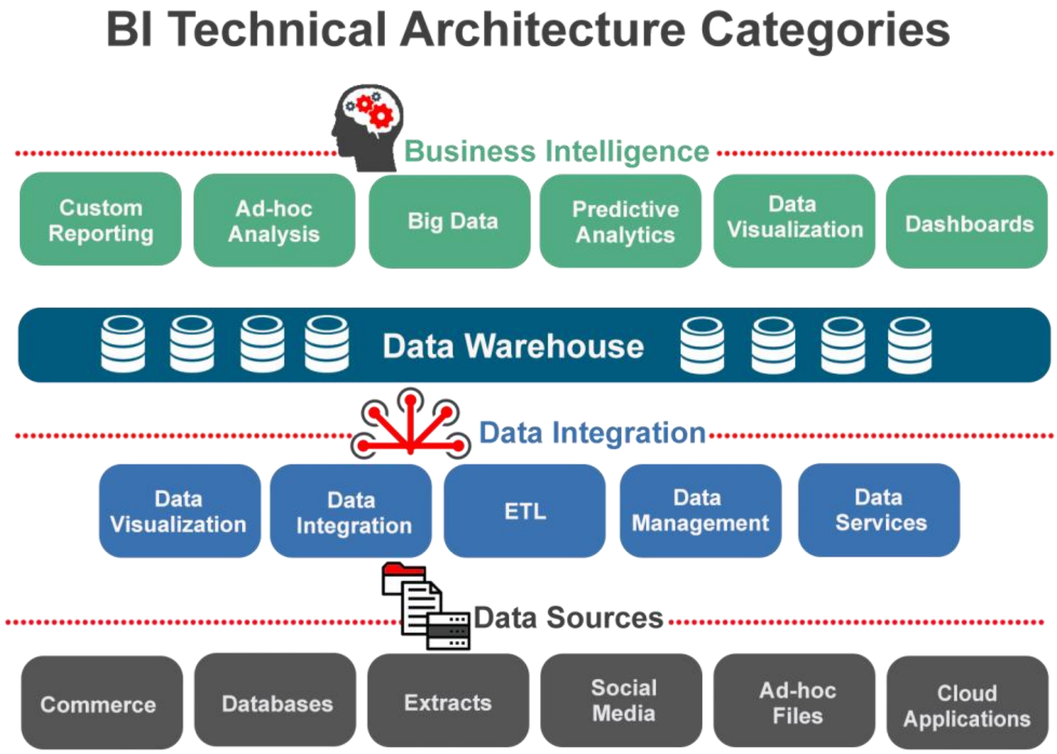


Fig. 9. Technical architecture categories of data

Data Processing Evolution:

Batch Processing → Stream / Real-time Processing → Lambda → Kappa /Delta

Initially, the big data solutions were typically long-running batch jobs to filter, aggregate, and otherwise prepare the data for analysis. Usually, these jobsinvolved reading source files from scalable storage like the Hadoop Distributed File System (HDFS), processing them, and writing the output to new files in scalable storage. The key requirement for these batch processing engines is the ability to scale out computations in order to handle a large volume of data. The stream, or real-time processing, deals with data streams captured in real-time and processed with minimal latency to generate real-time (or near-real-time) reports or automated responses. Frameworks like Apache Kafka, Apache Storm, Apache Spark Streaming, Amazon Kinesis, etc., help enable this capability.

The next evolution was the Lambda architecture, adata-processing architecturedesigned to handle massive quantities of data by taking advantage of both batch and stream-processing methods. This approach to architecture attempts to balance latency, throughput, and fault-tolerance by using batch processing to provide comprehensive and accurate views of batch data while simultaneously using real-time stream processing to provide views of online data Jay Kreps then proposed the Kappa architecture as an alternative to the Lambda architecture. It has the same primary goals as the Lambda architecture but with an important distinction: all data flows through a single path, using a stream processing system.

Another alternative is the Delta architecture, which introduces a table storage layer to handle stream and table storage accessed through a single code base. Databricks proposed this architecture, with Delta Lake at the center of the architecture. Delta Lake is an open-source atomicity, consistency, isolation, durability (ACID) table storage layer over cloud object stores. The Lambda architecture and its variants, the Kappa and Delta architecture, will continue to be valuable architectures in the near future. This concludes the first part of the blog series. We'll continue to explore the evolution of the data and analytics space in subsequent blog posts in this series in the coming months.

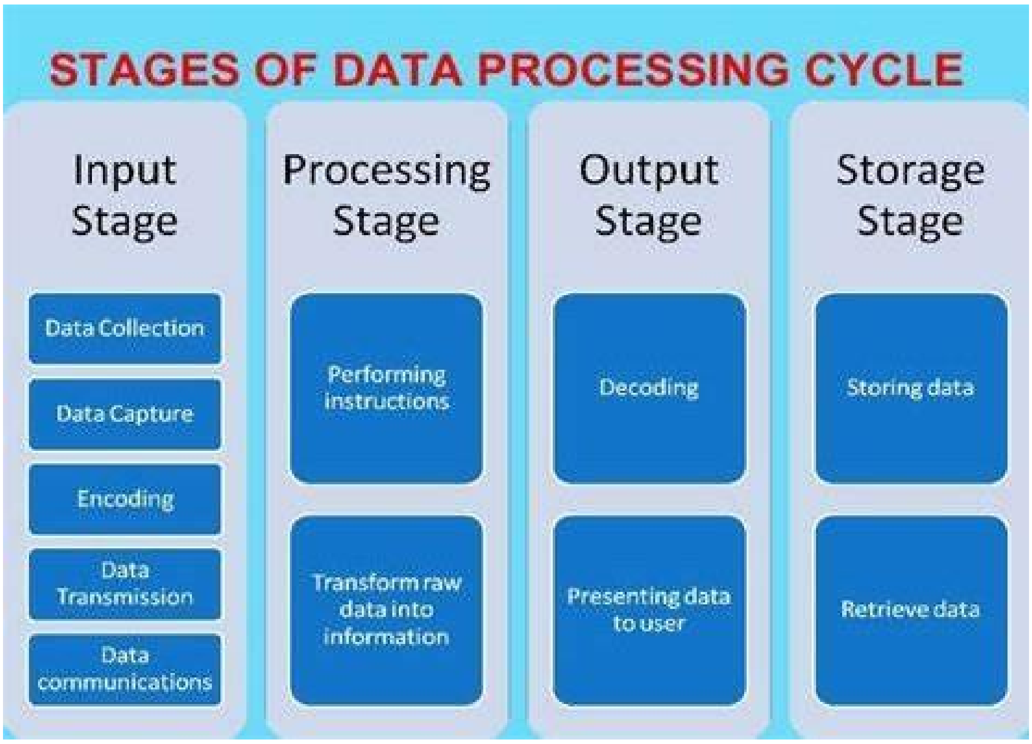


Fig. 10. Data processing cycles

The invention of computers created a clear need for information and data processing. During these very early days, computer scientists had to write custom programs for processing data and these were most likely stored on a punch card. The next steps brought assembly language and more purposeful programming languages like Fortran, followed by C and Java. During the prehistoric big data space, software engineers would use these languages to write purpose-built programs for specific data processing tasks. However, this data processing paradigm was only accessible to a select few that had a programming background which prevented wider adoption by data analysts or the wider business community who wanted to process information and make specific decisions.

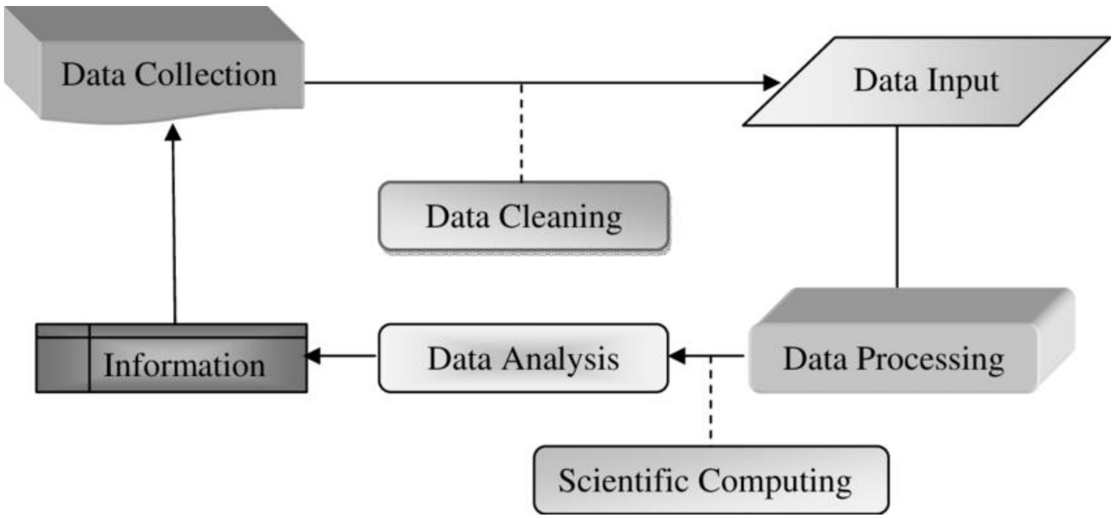


Fig. 11. Data processing steps

2.2. Data Analytics categories and Use Cases

Big Data Analytics is the ideal example of a contemporary technology and an industry disruption that can benefit every industry and every business organization. By 2023, the big data analytics market is expected to reach 103 billion dollars, and big data is already used in 70% of large enterprise business setups. Organizations continue to produce enormous amounts of data each year, and by 2025 it is predicted that the total amount of data produced, stored, and used worldwide will exceed 180 zettabytes.

Since many business leaders use data to make informed decisions, data analytics is significant across a wide range of industries. A sneaker manufacturer may use sales information to choose which styles to keep and which to retire, or a health care administrator may use inventory information to choose which medical supplies to order. Data on enrollment may be used by Coursera to decide what kinds of courses to add to its catalog. Whether they are aware of it or not, people use data every day and it is present everywhere. Analyzing and using data can be done in everyday activities like measuring coffee beans to make your morning cup, checking the weather before deciding what to wear, or tracking your steps with a fitness tracker.

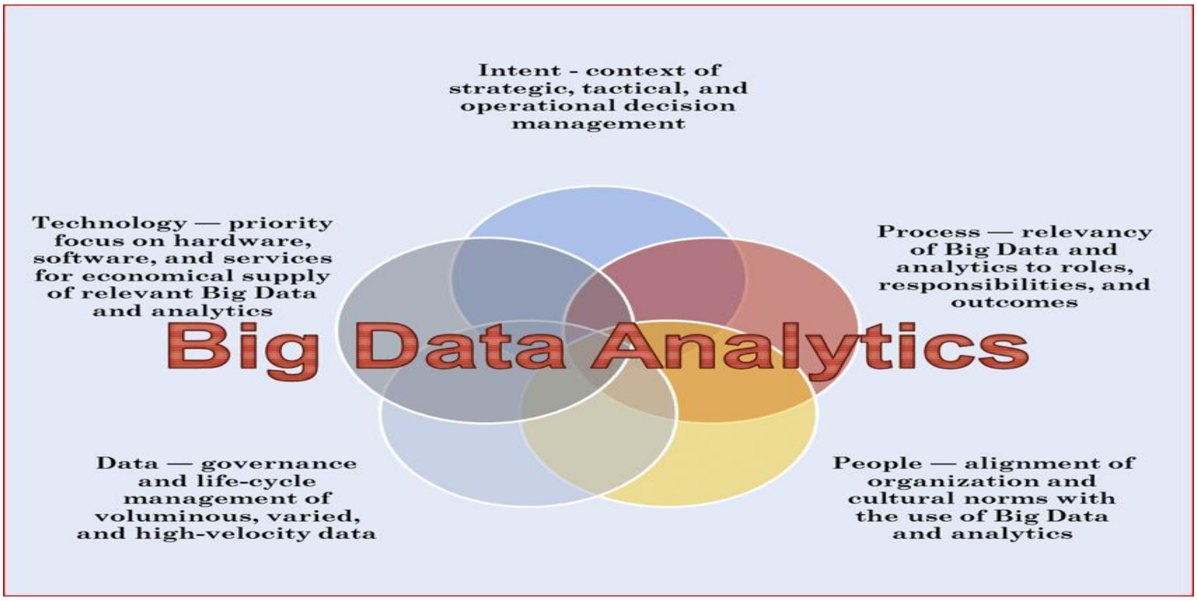


Fig. 12. Big data analytics

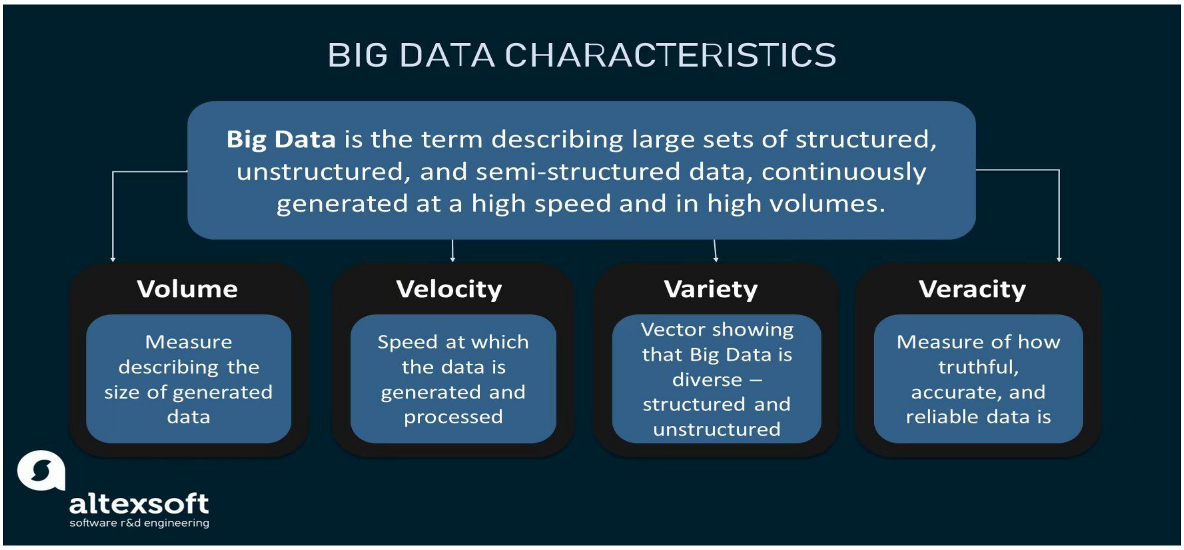


Fig. 13. Characterstics of Big Data

Data Analytics Use Cases in Various Industries Includes

1. Banking and Finance (Fraud Detection, Risk & Insurance, and Asset Management)

Futuristic banks and financial institutions are capitalizing on big data in various ways, ranging from capturing new markets and market opportunities to fraud reduction and investment risk management. These organizations are able to leverage big data analytics as a powerful solution to gain a competitive advantage as well.

Fig. 14. various use cases of big data analytics in the finance and banking sector

2. Accounting

Data is Accounting's heart and using big data analytics in accounting will certainly deliver more value to the accounting businesses. The accounting sector has various activities, such as different types of audits, checking and maintaining ledger, transaction management, taxation, financial planning, etc.

The auditors have to deal with numerous sorts of data that might be structured or unstructured, and big data analytics can help them in:

- Outliers identification
- Exclude exceptions
- Focus on data blocks of greatest risk areas
- Visualize data
- Connect financial and non-financial data
- Compare predicted outcomes for improving forecasting etc

3. Aviation

Studies reveal that the aviation analytics market will hit the 3bn USD by 2025 and will register a CAGR of 11.5% over the forecast period.

The major growth drivers of the aviation market are:

- Increasing demand for optimized business operations
- COVID-19 outbreak affecting the normal aviation operations
- Mergers, acquisitions, and joint ventures

Recent trends and modifications in the aviation industry's Original Equipment Manufacturer (OEM) and user segments Cloud-based real-time data collection and analytics, which call for a variety of data models, are one of the most lucrative big data analytics opportunities in the aviation sector. Similar to other industries, big data analytics has enormous potential in the airline sector as well, improving everything from routine tasks like maintenance, resource allocation, flight safety, and flight services to corporate objectives like loyalty programs and route optimization.

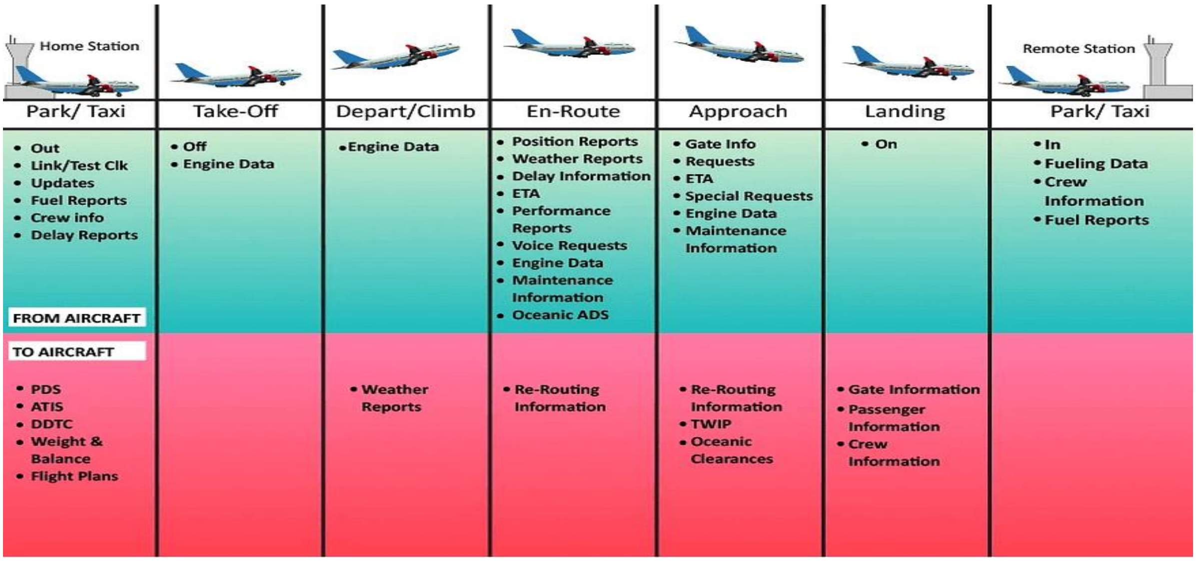


Fig. 15. the various points of data generation in the aviation industry (flights only), that can be a valid use case for big data analytics

4. Government and Law Enforcement

Government and public infrastructure produce a large amount of data in various forms, such as body cameras, CCTV footage, satellites, public schemes, registrations, certifications, social media, etc.

Big data analytics can empower the government and public services sector in many ways, some of which are mentioned below:

- Open data initiatives to manage, monitor, and track the private company data

- Encouraging public participation and transparency in open data initiatives by the government
- Predicting consumer frauds, political shifts, and tracking the border security
- Defense and consumer protection
- Public safety via a rapid and efficient address of public grievances
- Transportation and city infrastructure management
- Public health management
- Efficient and data-driven management of energy, environment, and public utilities

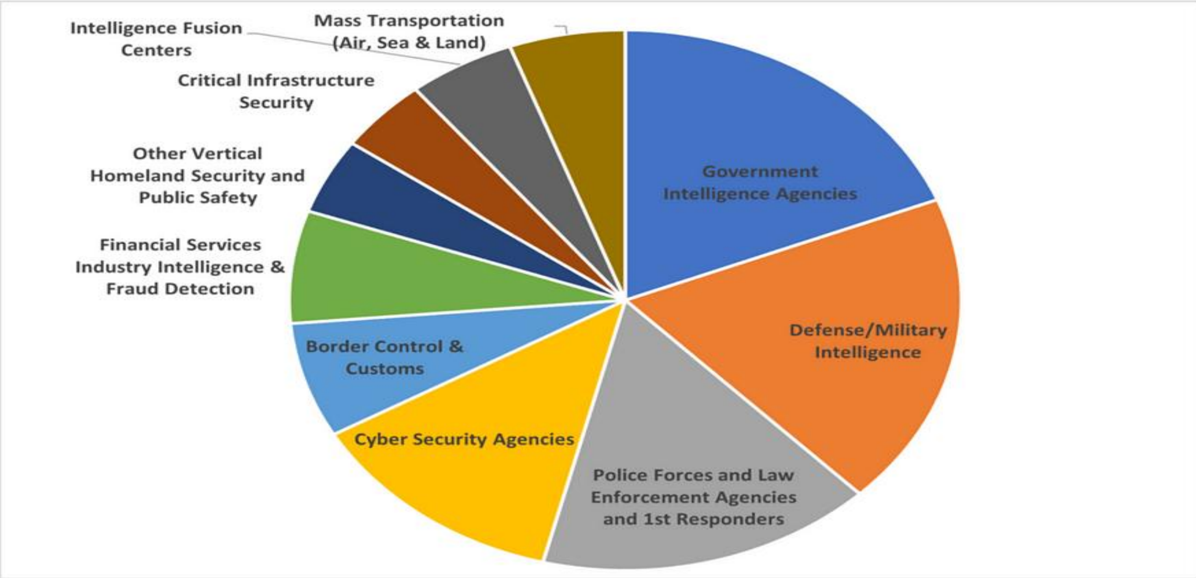


Fig. 16. shows how big data analytics can help the law enforcement and national security sectors

Big data analytics are also of utmost significance to the law enforcement sector. There are many ways big data analytics can assist law enforcement stakeholders, including tracking crimes, real-time and round-the-clock policing of sensitive areas, real-time monitoring and tracking of criminals, smugglers, and tracing money launderers.

2.3. The Needs for Big Data Analytics

It is difficult to store and retrieve enormous amounts of structured and unstructured data in a timely manner. The term "Big Data" first came into use as a result of some of these constraints on how much data could be handled and processed using conventional storage methods. Despite the fact that big data has gained attention as a result of the Internet's development, it cannot be compared to it. Beyond the Internet, however, the Web facilitates the gathering and exchange of knowledge as well as raw data. Big Data is the study of how to store, analyze, and comprehend data so that it can be used to predict future events with high accuracy and a tolerable time lag.

Big data analytics is currently and rapidly focusing on traditional methods like rule-based systems, pattern mining, decision trees, and other data mining techniques to efficiently develop business rules even on large data sets. Creating algorithms that make use of distributed data storage, in-memory computing, or cluster computing for parallel computation can all be used to achieve this. Grid computing, which was previously used for these processes but has recently been replaced by cloud computing, was used in the past.

2.3.1. Grid Computing

Grid computing is a method of distributing computing power to solve problems that are frequently huge and call for a lot of processing time and power. It operates under the voluntary basis principle, where users donate their memory and computing resources for use by others. In this situation, it's important to only use computers when necessary and scale the issues so even modest computers can contribute to the grid. Every computer that is connected to the Internet and desires to join the grid is regarded as a node in a very sizable computer.

- ✓ Commercial Model: It operates under the tenet that this technology can be used for commercial purposes by building sizable processing centers and renting out its capabilities to users for a fee on an hourly basis. This model has the benefit of maintaining Quality of Service (QoS), which is a reliable method of computation.
- ✓ Social Model: It is based on the idea that these resources ought to be used for the benefit of society. Through the use of software that adheres to the Open Grid Service Architecture (OGSA), the concept of grid computing is implemented. The widely used Globus toolkit implements OGSA and is used in grid computing environments.

Grid computing has a number of flaws, including problems with money, society, law, and regulations. The ownership of the results is another issue that is raised by distributing a commercial project across these donated machines. Given the zeal with which hackers have produced email viruses and video game cheats, these uninvited issues are connected to the advantages of grid computing.

2.3.2. Cloud Computing

Although the term "cloud computing" has recently gained popularity, the idea is not entirely new. Its foundations are found in utility computing, cluster computing, and distribution systems more broadly, as well as grid computing and other related fields. The term "cloud computing" refers to the idea of computing at a distance under user control using a thin client system, computer, or even a mobile device. The infrastructure of the service providers will house the processing, memory, and storage (Figure 17). Users must connect to the virtual system located at a distant location, which may use virtualization to run multiple virtual operating systems on physical servers. According to Bhadani and Chaudhary (2010), it supports a variety of fault-tolerant features like live migration, scalable storage, and load balancing.

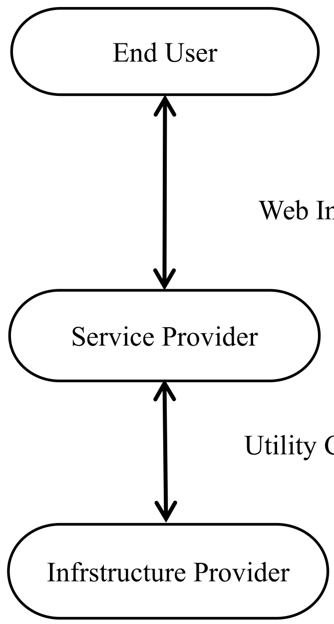


Fig. 17. Basic Model of Cloud Computing

However, given the advantages and disadvantages of grid computing or cloud computing, it might be useful to process the enormous amount of data that must be processed for big data analysis or live stream data analysis (Foster et al., 2008). No technology is completely foolproof.

2.4. The use of data analytics in many businesses

Data analytics allows a business to create reports and find patterns that can help it operate more efficiently. Analyzing data can also improve decision-making by enabling a company topredict trends in its industry or predict what its customers will want next. These predictions help companies stay on the cutting edge and remain competitive.

For example,a pizza shop may use analytics to understand the demographics of its customers or how much they spend at the restaurant. This can help them plan promotions and marketing campaigns better. Data analytics can be helpful to understand the different types of customers who visit the store. For example, if the pizza shop discovers that most of its customers are families with children, it may want to offer more kid-friendly items. On the other hand, if students are regular customers, they may like to provide student discounts as marketing incentives. The pizza shop might also use data analytics to assess employee performance based on sales data from each server. If one server has low sales throughout the day, his boss may want to check on him and see if he needs more training or is not performing up to standards.

In the era of big data, where business leaders have access to more information than ever before, analyzing this information is an important skill for any professional in a leadership role. That's why many employers make business analytics training a priority for their employees who are looking to advance within the company.

Data analytics use cases in business:

The use of data analytics in business is becoming increasingly important. Here is a look at some common data analytics use cases adopted in various business functions.

i). Data analytics use cases in marketing & sales: Marketing and sales are the most popular areas where companies have implemented data analytics use cases. Both sectors benefit from the use of data in separate ways. With the help of big data,marketing managers can now make more accurate decisions.They collect data from various sources, such as social media, email communication, and other platforms, to communicate with their customers to understand their needs and preferences better. Marketing specialists then analyze the collected data to improve the marketing strategy and drive more sales.

Data analytics can help salespeople sell more effectively.For example, with the help of data analytics, sales representatives can easily understand which products are most demanded by particular target groups and focus on selling them to increase their performance level.In marketing and sales, data analytics can help businesses to:

- Identify which customers are most likely to respond to a specific offer
- Find the best product mix for a given customer
- Find the best price for a given product or product mix
- Determine the effectiveness of different advertising or marketing campaigns
- Identify new markets for existing products or services

ii). Data analytics use cases in operations and supply chain management:

Any business with a supply chain or operations component can benefit from data analytics. For example, suppose you manufacture widgets, and those widgets require parts made by suppliers. In that case, you want to anticipate when your inventory is getting low so that you don't run out of parts before they can be delivered. You might also use data analytics to optimize production schedules and staffing levels. Here are some of the data analytics use cases in operations and supply chain management:

Supplier selection and performance evaluation: Usingmachine learning algorithms, supplier selection processes can be improved dramatically. In addition, companies can use data analytics to predict and manage supplier defaults.

Predictive maintenance: Predictive maintenance techniques enable companies to collect and analyze data from sensors to understand the state of equipment better and predict failures before they occur.

Location optimization: Companies can use location-based analytics solutions such as GIS (geographic information systems) to determine optimal locations for warehouses, factories, and service centers.

*Inventory management:*Data analytics can help companies better understand customer demand patterns to improve inventory control and reduce inventory costs.

iii). Data analytics use cases in human resources:

The use cases for data analytics in business are diverse, but one area ripe with opportunity is human resources. A study by a popular research agency found that 84 percent of HR executives surveyed invest in analytics to improve talent management. Of the 716 companies surveyed, over half (56 percent) plan to increase analytics spending in the next two years.

Those investments are paying off: 63 percent of the companies surveyed said they already have a return on their investment in analytics.

iv). Data analytics use cases in customer service:

The customer is king, and most companies want to understand their customers' wants. Using data analytics and artificial intelligence, they can nowget deeper insights into customer behavior. Most businesses use data analytics in customer service in one form or another. It may be as basic as tracking metrics like number of calls or wait time on hold. But companies can also use data analytics to understand better customers' needs, preferences, satisfaction levels, and more.

In the near future, you'll likely see more companies take advantage of these types of data use cases:

- Identify common complaints and problems that customers have about products or services
- Resolve issues faster and more effectively through a better understanding of customer history and needs
- Predict what products or services a customer is likely to buy next
- Provide customers with personalized content and recommendations
- Automate processes such as payment processing and fraud detection
- Reduce the cost of delivering support, such as by providing self-service options

v. Data analytics use cases in finance:

Finance is one of those departments that seems like it would have been made For big data after all, financial data has been tracked by computers since there were computers. But actually, it's quite likely that yourfinance department is still doing most of its work manually.And this means that it's missing out on all sorts of opportunities for efficiency and innovation.

Here are some use cases of data analytics in finance departments:

- Analyze and predict financial performance, such as sales trends and profit margins
- Measure the efficiency of marketing campaigns and make informed decisions on where to put more money into and which one to cut back
- Optimize pricing strategies based on market conditions and customer demand
- Reduce fraudulent activities by identifying unusual patterns, which could be a sign of a fraud
- Forecast spending and revenue
- Predict cash flow, manage budgets, and determine cash requirements

2.5. Data sources and tools for processing data

Understanding what the data contains and does not contain depends on how the data are organized and thought about. There are many different approaches to data analysis, and it is infamously simple to manipulate data during the analysis phase in order to push particular conclusions or agendas. There is a growing need for quicker and more effective methods of analyzing this data due to the development of technology and the vast amounts of data that are now flowing into and out of organizations every day. Making effective decisions at the appropriate time requires more than just having a ton of data available. Traditional data management and analysis methods and infrastructures are no longer suitable for the quick analysis of such data sets.

As a result, new big data analytics-specific tools and techniques as well as the necessary architectural frameworks for storing and managing such data are needed. As a result, everything from the data itself and its collection to its processing to the final extracted decisions is impacted by the emergence of big data. In order to incorporate big data analytics tools and methods into the decision-making process, the Big Data, Analytics, and Decisions (B-DAD) framework was proposed. The framework links various big data management, processing, and analysis tools, as well as visualization and evaluation tools, to the various stages of the decision-making process.

Hence, the changes associated with big data analytics are reflected in three main areas: big data storage and architecture, data and analytics processing, and, finally, the big data analyses which can be applied for knowledge discovery and informed decision making. Each area will be further discussed in this section. However, since big data is still evolving as an important field of research, and new findings and tools are constantly developing, this section is not exhaustive of all the possibilities, and focuses on providing a general idea, rather than a list of all potential opportunities and technologies.

There are mostly two kinds of origins of information or data sources, Researchers use both data sources a lot in their work. The data is collected from these using either primary or secondary research methods.

Types of data sources

1. Statistical data sources

Surveys and other statistical reports used for official purposes are sources of statistical data. People are asked a number of questions both qualitative and quantitative here. Quantitative data uses numbers, whereas qualitative data sources do not. Both types of statistical data are used in the data sampling technique. A statistical survey is typically conducted using a sample of respondents. This approach involves gathering sample data, which is then put through a statistical analysis. The questionnaire approach can also be used to conduct the surveys.

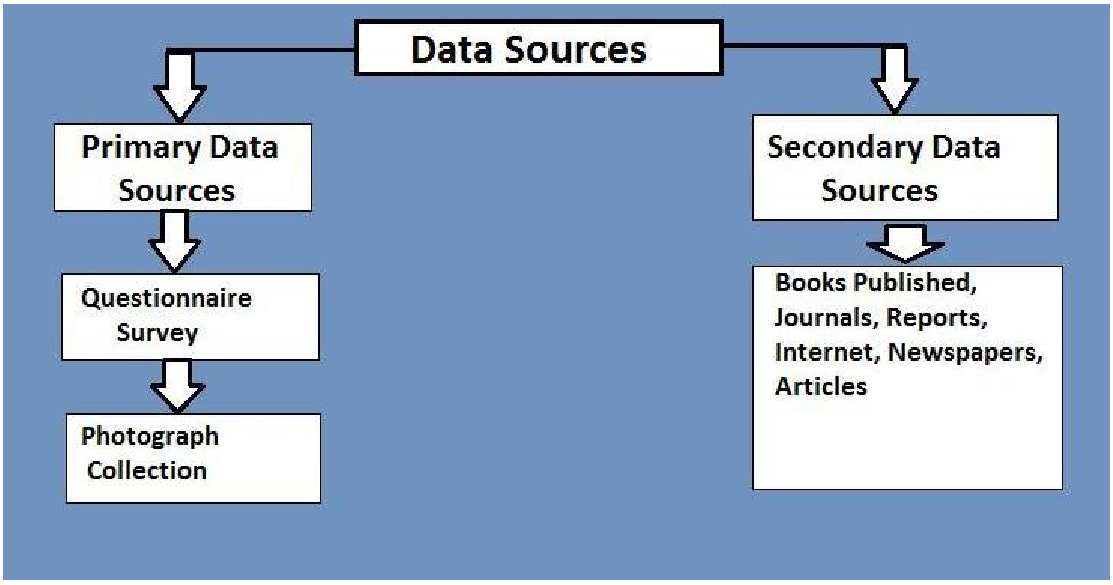


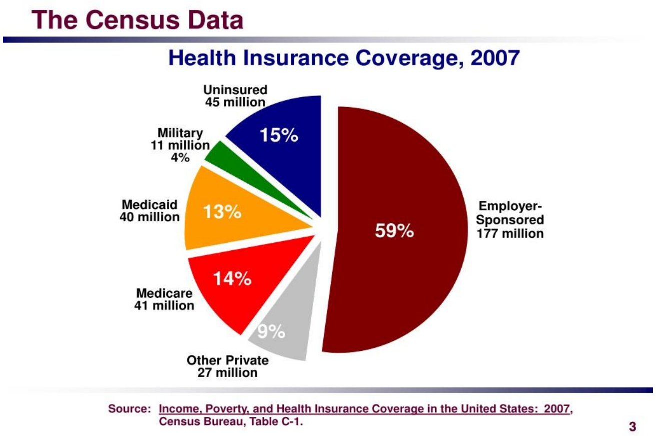
Fig. 18. Statistical data source

2. Census data sources

The information used in this method comes from a census report that was previously released. In contrast to statistical surveys, it. Throughout the research process, the Census method closely examines every component of the population. Here, data is gathered over a predetermined period of time known as the reference time. The researchers conduct their research at a specific time, analyze it, and draw conclusions. The nation conducts censuses for official purposes. Questions are posed to the respondents, who then respond. This conversation can happen over the phone or in person. However, because it involves the entire population, the census is a data source that requires a lot of time and effort.

While "the census" refers to the population count that occurs every ten years, the US Census Bureau actually regularly publishes a wide range of datasets. These geographically and thematically detailed free and open datasets can be used to research a wide range of housing, socioeconomic, and demographic trends. This guide gives you a general overview of census ideas and datasets and directs you to tools and resources you can use to access and extract data.

In addition, the census, a regular and systematic record of population data, is one of the many ways a government can gather population data. In order to produce the most accurate results, a census is designed to collect information from as many people as is physically feasible. This indicates that they are very significant. All adult males were counted in ancient Rome's census, which gave the emperor information he could use to determine how many men to conscript into an army if necessary. For you Bible enthusiasts, the census was the reason Joseph and Mary were going to Jerusalem. Nevertheless, the modern census counts everyone, not just men, and even though it is primarily done digitally now, for many years official census takers would travel the world and manually count people.



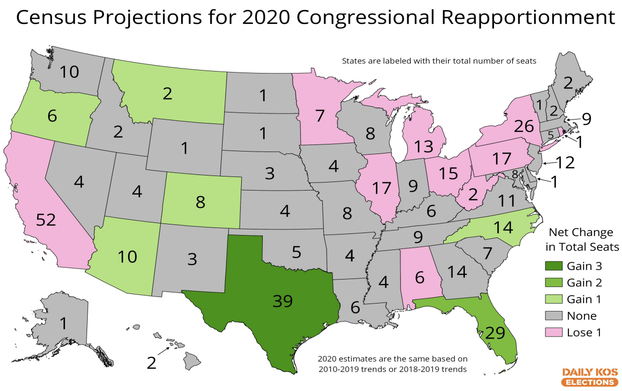


Fig 19. Census data sources of health insurance coverage and the 2020 congressional reapportionment

3. Methodology

Research methodology and design play a crucial role in the accuracy and validity of research findings. This section discusses the research methodology and design employed in this study on advances in data analytics technologies and applications.

The research will use a qualitative research methodology, which involves gathering data from various sources such as academic journals, online databases, and relevant websites. The data will be analyzed using thematic analysis, which involves identifying patterns, trends, and themes in the data. The research will consider ethical considerations and limitations to ensure that the data is gathered ethically and is valid.

3.1. Research methodology and design

The primary objective of this research is to explore and evaluate recent developments in data analytics technologies and their applications across various industries. To achieve this objective, a mixed-methods research design was utilized, combining both qualitative and quantitative data collection and analysis techniques.

The research began with a comprehensive review of relevant literature, which provided a solid foundation for the study. The review involved an overview of data analytics technologies and their evolution, data analytics categories and use cases, data sources and tools for processing data, technologies and problems associated with big data analytics, artificial intelligence and machine learning in data analytics, data visualization and analytics tools, the use of data analytics in businesses, data analytics technologies' challenges, and future directions.

The qualitative research component of this study involved a series of semi-structured interviews with experts in the field of data analytics technologies. The participants were selected based on their expertise and experience in the development and implementation of data analytics technologies across different industries. The interviews were conducted face-to-face, and a set of pre-defined questions were used to guide the discussion. The interviews were audio-recorded, transcribed, and analyzed using a thematic analysis approach.

To ensure the validity and reliability of the research findings, various measures were taken. Firstly, the research questions were explicitly stated, and the research design was chosen to address these questions. Secondly, the sample sizes for both the interviews and the survey were chosen based on the research population and were deemed adequate for the research questions. Thirdly, the research instruments were pre-tested to identify any potential errors or issues. Fourthly, the data collected were analyzed using appropriate statistical and qualitative analysis techniques. Finally, the research findings were discussed in the context of the existing literature, and conclusions were drawn based on the research objectives.

3.2. Data collection and sources

Data collection is a fundamental aspect of data analytics technologies, particularly in the context of advanced data analytics that rely on large, diverse, and complex data sources. Data collection involves gathering, storing, and organizing data from various sources to enable its analysis and interpretation. Advanced data analytics technologies require access to high-quality, reliable, and relevant data to generate insights, predictions, and recommendations. The choice of data sources and collection methods can significantly affect the outcomes of data analytics research and applications.

Data Sources for Advanced Data Analytics Technologies:

Data sources for advanced data analytics technologies can be categorized into three types: structured, semi-structured, and unstructured data. Structured data refers to data that is organized and formatted in a predefined manner, such as relational databases, spreadsheets, and data warehouses. Semi-structured data refers to data that has some structure but is not fully organized or standardized, such as XML and JSON files. Unstructured data refers to data that has no structure or organization, such as text documents, images, videos, and social media feeds. Advanced data analytics technologies can leverage all three types of data sources to generate insights and predictions.

Traditional Data Collection Methods:

Traditional data collection methods for advanced data analytics technologies include surveys, interviews, experiments, and observations. Surveys involve asking questions to a sample of respondents to gather data on their attitudes, opinions, and behaviors. Interviews involve one-on-one or group discussions with participants to gather data on their experiences, perspectives, and attitudes. Experiments involve manipulating variables in a controlled environment to test hypotheses and gather data. Observations involve systematically observing and recording behavior or phenomena to gather data. Traditional data collection methods are useful in gathering high-quality, structured data but can be time-consuming, expensive, and prone to biases and errors.

Emerging Data Collection Methods:

Emerging data collection methods for advanced data analytics technologies include crowdsourcing, sensor networks, and social media data mining. Crowdsourcing involves engaging a large number of people through online platforms to gather data, such as opinions, preferences, and feedback.

Sensor networks involve deploying sensors, which are capable of collecting data from their surrounding environments, to monitor physical or environmental conditions. The sensors collect data on temperature, humidity, pressure, and other physical parameters, and then transmit this data to a central data collection point, where it can be processed and analyzed using data analytics tools.

One of the advantages of using sensor networks for data collection is that it enables the monitoring of conditions in real-time. This allows for faster detection of changes in the environment, which can be crucial in situations where quick action is necessary. Additionally, sensor networks can be deployed in a variety of settings, from urban environments to remote locations, and can provide data on a wide range of physical parameters.

There are several challenges associated with data collection using sensor networks, however. One major challenge is the sheer amount of data that is generated by sensor networks. In some cases, the data generated can be so massive that it is difficult to process and analyze using traditional data analytics techniques. This has led to the development of new data processing and analysis techniques specifically tailored to sensor networks.

Another challenge associated with data collection using sensor networks is the issue of data quality. Sensors can be prone to errors or inaccuracies, and data collected from sensors may need to be filtered or processed in order to remove errors or other forms of noise. This can be a time-consuming process, and requires careful attention to detail in order to ensure that the data collected is of high quality.

Despite these challenges, the use of sensor networks for data collection continues to grow in popularity. Advances in technology have made it easier and more cost-effective to deploy sensor networks, and new data processing and analysis techniques are being developed to handle the large amounts of data generated by these networks.

As a result, the use of sensor networks is expected to continue to expand in a wide range of industries and applications.

Data Analysis Techniques

Cluster Analysis:

Cluster analysis is a technique that is used to group data points with similar characteristics. The analysis involves identifying groups of data points that are similar to each other based on a specific set of variables. The clusters are identified using mathematical algorithms that analyze the similarities and differences between data points. Cluster analysis is widely used in market research to identify groups of consumers with similar characteristics. It can also be used in healthcare to group patients with similar symptoms, or in biology to group species with similar characteristics.

Factor Analysis:

Factor analysis is a statistical technique that is used to identify the underlying factors that contribute to a set of observed variables. The technique involves analyzing the correlation between different variables to identify which variables are strongly correlated and which are not. Factor analysis is commonly used in social science research to identify the underlying factors that contribute to a specific behavior or trait. For example, factor analysis could be used to identify the factors that contribute to job satisfaction or employee turnover.

Time Series Analysis:

Time series analysis is a statistical technique that is used to analyze data points collected over time. The technique involves analyzing the patterns, trends, and cycles in the data to identify any significant changes or patterns.

Time series analysis is widely used in finance to analyze stock prices, in economics to analyze trends in economic indicators, and in meteorology to analyze weather patterns. It can also be used in healthcare to analyze patient data collected over time to identify any patterns or trends in their symptoms or health outcomes.

3.3. Data analysis and interpretation

Data Interpretation

Data interpretation is a critical stage in the data analysis process that involves drawing insights and conclusions from the analyzed data. It is an iterative process that involves using analytical techniques and domain knowledge to make sense of the data and derive meaningful insights. Interpretation involves understanding the patterns, trends, and relationships in the data, as well as the implications of the findings.

One of the key considerations in data interpretation is understanding the context in which the data was collected. This includes understanding the data sources, the data collection methods, and any limitations or biases that may be present in the data. Understanding the context of the data is crucial for ensuring that the conclusions drawn from the analysis are accurate and relevant.

Another important consideration in data interpretation is the selection of appropriate visualization techniques to communicate the findings effectively. Visualization techniques can include graphs, charts, tables, and other graphical representations that can help to convey the meaning of the data in a clear and concise manner. Effective visualization techniques can make it easier for stakeholders to understand the results and take appropriate action based on the findings.

Data interpretation also involves identifying and addressing any gaps or inconsistencies in the data. This can involve further analysis or data collection to fill in the gaps, as well as verifying the accuracy of the data through cross-validation and other validation techniques. Addressing gaps and inconsistencies in the data is important for ensuring that the conclusions drawn from the analysis are robust and reliable.

Finally, data interpretation involves making recommendations based on the insights obtained from the analysis. These recommendations may involve changes to business processes, products, or services, or other actions that can help to improve performance, reduce costs, or achieve other desired outcomes. It is important to ensure that the recommendations are actionable and relevant to the stakeholders who will be implementing them.

Overall, data interpretation is a critical component of the data analysis process that involves drawing insights and conclusions from the analyzed data. It requires a combination of analytical techniques, domain knowledge, and communication skills to effectively communicate the meaning of the data to stakeholders. By following best practices in data interpretation, organizations can leverage the power of data to drive business success.

3.4. Limitations and ethical considerations

The methodology section of a research project should also include a discussion of the limitations of the study and any ethical considerations that were taken into account. Limitations refer to any factors that may affect the validity or reliability of the results, while ethical considerations refer to any ethical issues that arise during the research process.

A. Limitations

Limitations are an important aspect of any research study, as they define the boundaries within which the results of the study can be applied. In other words, limitations are the factors that may affect the validity or reliability of the results, and which should be taken into account when interpreting the findings of the study.

Limitations can be either internal or external to the study, and can arise from a variety of factors.

Internal limitations refer to factors that are within the control of the researcher, such as the research design, sampling methods, data collection instruments, and data analysis techniques. One common internal limitation is the use of a small sample size, which can limit the generalizability of the findings. For example, a study that only includes a few participants from a specific population may not be representative of the entire population, and the results may not be applicable to other settings or populations. Other internal limitations may include issues with the reliability or validity of the data collection instruments, such as survey or interview questions that are poorly designed or difficult to understand.

External limitations refer to factors that are outside the control of the researcher, such as the context in which the study is conducted, the characteristics of the study population, and other external factors that may influence the results. For example, a study that is conducted in a specific geographic area may not be applicable to other regions with different cultural or socio-economic factors. Similarly, the characteristics of the study population, such as age, gender, or education level, may limit the generalizability of the findings to other populations with different characteristics.

It is important for researchers to acknowledge and address the limitations of their study in order to provide a clear understanding of the strengths and weaknesses of the research. This can be done by including a discussion of the limitations in the methodology section of the research paper or report. The discussion should clearly outline the limitations of the study, and explain how these limitations may have affected the results. Additionally, researchers may offer suggestions for future research that can help to overcome these limitations and improve the validity and reliability of the results.

Ethical considerations are also an important aspect of any research study, and refer to the ethical issues that arise during the research process. Ethical considerations may include issues related to informed consent, privacy and confidentiality, and the potential for harm to study participants. Researchers must ensure that they have obtained informed consent from all participants, and that the data collected is kept confidential and secure. Additionally, researchers must take steps to minimize any potential harm or discomfort that may be experienced by study participants.

In summary, limitations and ethical considerations are important factors that should be taken into account when conducting any research study. These factors can affect the validity and reliability of the results, and must be acknowledged and addressed in the methodology section of the research paper or report. By acknowledging these limitations and ethical considerations, researchers can provide a clear understanding of the strengths and weaknesses of their study, and help to ensure that their research is conducted in an ethical and responsible manner.

B. Ethical Considerations

Certainly, ethical considerations play a crucial role in any research project, and it is important to take into account any ethical issues that may arise during the research process. The following is a discussion of some of the key ethical considerations that researchers should take into account:

Informed Consent: Informed consent is a fundamental ethical principle in research, and it involves obtaining the voluntary consent of participants to participate in the study. Informed consent means that participants are fully informed about the purpose of the study, the procedures involved, any risks or benefits, and their right to withdraw at any time without penalty. Researchers should ensure that participants are fully informed and have the capacity to provide informed consent before they are included in the study.

Confidentiality and Privacy: Confidentiality and privacy are important ethical considerations in research, particularly when dealing with sensitive information. Researchers should take steps to ensure that participants' personal information is kept confidential and that their privacy is respected. This may involve using pseudonyms, coding data, or limiting access to data to only authorized personnel.

Risks and Benefits: Researchers should consider the risks and benefits of the research to participants, as well as to society as a whole. Participants should not be exposed to undue risks, and any potential benefits of the research should outweigh any potential risks. Researchers should also consider the potential impact of the research on society, particularly if the research has the potential to result in harm.

Deception: Deception is sometimes necessary in research, particularly in social science research, but it must be kept to a minimum. Researchers should ensure that any deception used in the study is justified, and that participants are debriefed as soon as possible after the study to minimize any negative effects.

Coercion: Coercion involves using pressure or force to make someone participate in the study. Researchers should ensure that participants are not coerced into participating in the study and that they have the right to withdraw at any time without penalty.

Respect for Participants: Researchers should respect the autonomy and dignity of participants and treat them with respect. This involves taking steps to ensure that participants are not exploited or treated unfairly, and that their cultural and social backgrounds are taken into account.

Conflict of Interest: Researchers should be transparent about any potential conflicts of interest that may arise during the study. This includes any financial or personal interests that may influence the research results.

It is important to note that ethical considerations can vary depending on the nature of the research project, the participants involved, and the cultural and social context in which the research is conducted. Therefore, researchers should always be vigilant and take steps to ensure that their research is conducted in an ethical manner.

4. Results and Insights

The results of this report reveal that there have been significant advancements in data analytics technologies. These advancements have enabled businesses to analyze large amounts of data quickly and efficiently. Some of the key technologies that have been developed include machine learning algorithms, natural language processing, and deep learning. These technologies have been applied in various industries to optimize processes, improve customer experiences, and develop new products and services.

4.1. A Summary of Recent Developments In Data Analytics Technology.

Data analytics technology has evolved rapidly in recent years due to the increasing demand for data-driven insights across various industries. Several new advancements have contributed to the development of data analytics technology. Artificial intelligence and machine learning algorithms are used for predictive modeling, data processing automation, and improving data accuracy. Big data platforms like Hadoop and Spark have become essential for managing and processing large datasets. Cloud-based analytics platforms offer scalability, flexibility, and cost-effectiveness. Natural language processing (NLP) techniques have been used to extract insights from unstructured data. IoT analytics tools have been developed for analyzing data from connected devices. Data visualization tools have become more sophisticated, allowing for easy understanding of complex data through interactive dashboards and visualizations. Edge analytics involves processing data closer to the data source, improving real-time data analysis and reducing data transmission costs. These recent advancements have led to more efficient data processing, improved accuracy, and better decision-making, and further advancements in data analytics technology can be expected as data continues to play a critical role in business and society.

AI and machine learning: Artificial intelligence (AI) and machine learning (ML) are becoming increasingly important in data analytics. ML algorithms are being used to analyze large data sets and identify patterns and trends that would be difficult or impossible to detect with traditional statistical methods. AI-powered chatbots and virtual assistants are also being used in data analytics to help businesses automate tasks and gain insights from unstructured data.

Cloud computing: Cloud computing has made it easier and more cost-effective to store, process, and analyze large amounts of data. Cloud providers like Amazon Web Services (AWS), Microsoft Azure, and Google Cloud Platform offer a wide range of tools and services for data analytics, including data storage, processing, and visualization.

Big data technologies: big data technologies like Hadoop, Spark, and NoSQL databases have revolutionized the way organizations handle large volumes of data. These technologies make it possible to store and process massive data sets quickly and efficiently, enabling businesses to gain insights that were previously impossible to obtain.

Edge computing: Edge computing is a new paradigm that brings computation and data storage closer to the devices and sensors that generate the data. By processing data locally, at the edge of the network, organizations can reduce latency and bandwidth requirements, while also improving data privacy and security.

Natural language processing (NLP): Natural language processing is a subfield of AI that focuses on the interaction between computers and humans using natural language. NLP technologies are being used in data analytics to help businesses analyze unstructured data such as customer feedback, social media posts, and online reviews.

Graph analytics: Graph analytics is a type of analysis that focuses on the relationships between entities in a data set. Graph databases and algorithms can be used to uncover patterns and connections between data points that would be difficult to detect with traditional relational databases.

Explainable AI (XAI): Explainable AI is a subset of AI that aims to make machine learning models more transparent and interpretable. XAI techniques help data analysts and decision-makers understand how AI models arrive at their predictions, making it easier to identify and correct errors or biases.

Quantum computing: Quantum computing is an emerging technology that has the potential to revolutionize data analytics. Quantum computers can perform certain types of calculations exponentially faster than traditional computers, which could enable organizations to solve complex data analysis problems more quickly and efficiently.

Augmented analytics: Augmented analytics is an approach that combines AI and ML with traditional analytics techniques to automate data preparation, insight discovery, and data visualization. Augmented analytics tools use natural language processing and machine learning to help non-technical users generate insights from data more easily.

Blockchain: Blockchain technology is being used in data analytics to help organizations verify the authenticity of data and prevent fraud. By creating a decentralized, tamper-proof ledger of all data transactions, blockchain technology can increase transparency and trust in the data analytics process.

4.2. Data Analysis Methods

Some professionals use the terms “data analysis methods” and “data analysis techniques” interchangeably. To further complicate matters, sometimes people throw in the previously discussed “data analysis types” into the fray as well! Our hope here is to establish a distinction between what kinds of data analysis exist, and the various ways it’s used.

Although there are many data analysis methods available, they all fall into one of two primary types: qualitative analysis and quantitative analysis.

4.2.1. Qualitative Data Analysis

The qualitative data analysis method derives data via words, symbols, pictures, and observations. This method doesn't use statistics. The most common qualitative methods include:

- ✓ Content Analysis, for analyzing behavioral and verbal data.
- ✓ Narrative Analysis, for working with data culled from interviews, diaries, surveys.
- ✓ Grounded Theory, for developing causal explanations of a given event by studying and extrapolating from one or more past cases.

4.2.2. Quantitative Data Analysis

Statistical data analysis methods collect raw data and process it into numerical data. Quantitative analysis methods include:

- ✓ Hypothesis Testing, for assessing the truth of a given hypothesis or theory for a data set or demographic.
- ✓ Mean, or average determines a subject's overall trend by dividing the sum of a list of numbers by the number of items on the list.
- ✓ Sample Size Determination uses a small sample taken from a larger group of people and analyzed. The results gained are considered representative of the entire body.

4.3. Data Analytics Tools

These data analysis tools are primarily concerned with making the lives of analysts easier by offering them solutions that streamline challenging analytical tasks. They have more time to complete the analytical portion of their work in this way. Now that we looked at the different steps involved in data analytics, let's see the tools involved in data analytics, to perform the above steps. In this blog, we will discuss 7 data analytics tools, including a couple of programming languages that can help you perform analytics better.

Fig 20.Data Analytics for Beginners - Tools used

1. Python: Python is an object-oriented open-source programming language. It supports a range of libraries for data manipulation, data visualization, and data modeling.

2. R: R is an open-source programming language majorly used for numerical and statistical analysis. It provides a range of libraries for data analysis and visualization.

3. Tableau: It is a simplified data visualization and analytics tool. This helps you create a variety of visualizations to present the data interactively, build reports, and dashboards to showcase insights and trends.

4. Power BI: Power BI is a business intelligence tool that has an easy 'drag and drop functionality. It supports multiple data sources with features that visually appeal to data. Power BI supports features that help you ask questions to your data and get immediate insights.

5. QlikView: QlikView offers interactive analytics with in-memory storage technology to analyze vast volumes of data and use data discoveries to support decision making. It provides social data discovery and interactive guided analytics. It can manipulate colossal data sets instantly with accuracy.

6. Apache Spark: Apache Spark is an open-source data analytics engine that processes data in real-time and carries out sophisticated analytics using SQL queries and machine learning algorithms.

7. SAS: SAS is a statistical analysis software that can help you perform analytics, visualize data, write SQL queries, perform statistical analysis, and build machine learning models to make future predictions.

Now that you have seen the data analytics tools, let's jump ahead and see the applications of data analytics. Unless you've been living in a cave for the past couple of decades, you know how important data has become virtually every business and every consumer on the planet. How much data is out there? A recent calculation estimates that the entire digital universe will reach 44 zettabytes in 2020. For perspective, that's "40 times more bytes than there are stars in the observable universe."

The key, of course, is knowing how to mine that data, analyze it, extract value from it, and apply it to a tangible business solution. Companies across industries and government organizations are using big data analytics and data science technologies to change the way they operate and to create solutions that impact people in almost every conceivable way.

According to a study by Accenture, 79 percent of executives feel that companies not embracing data analytics will lose their competitive position, and 83 percent have pursued big data projects to seize competitive advantage. The stories they are now telling about their successes are exciting. Here are four industries that are using advanced data analytics to thrive in a data-driven world, and important skills you'll need to be a part of it.

4.4. Data Analytics Applications

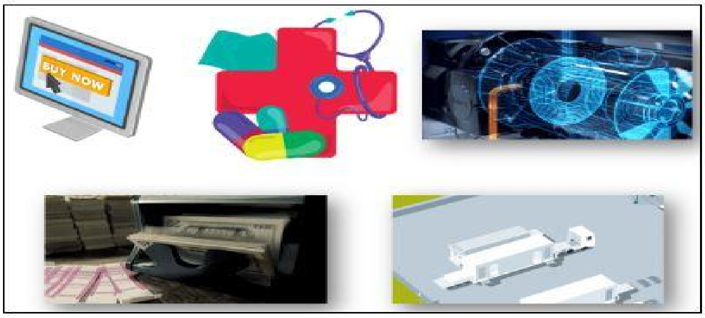


Fig. 21.Various applications of data analytics

Data analytics is used in almost every sector of business, let's discuss a few of them:

1. Retail: Data analytics helps retailers understand their customer needs and buying habits to predict trends, recommend new products, and boost their business. They optimize the supply chain, and retail operations at every step of the customer journey.

2. Healthcare: Healthcare industries analyze patient data to provide lifesaving diagnoses and treatment options. Data analytics help in discovering new drug development methods as well.

3. Manufacturing: Using data analytics, manufacturing sectors can discover new cost-saving opportunities. They can solve complex supply chain issues, labor constraints, and equipment breakdowns.

4. Banking sector: Banking and financial institutions use analytics to find out probable loan defaulters and customer churn out rate. It also helps in detecting fraudulent transactions immediately.

5. Logistics: Logistics companies use data analytics to develop new business models and optimize routes. This, in turn, ensures that the delivery reaches on time in a cost-efficient manner.

Those were a few of the applications involving data analytics. To make things simpler, this blog will also focus on a case study from Walmart. Here you can observe how data analytics is applied to grow a business and serve its customers better.

4.5. Walmart Case Study

The American multinational retail company- Walmart has over 11,500 stores in 27 countries worldwide. It also has e-commerce websites in 10 different countries. Walmart boasts more than 5,900 retail units. These units operate outside the United States, with 55 banners in 26 countries. It has more than 700,000 associates serving more than 100 million customers every week. In short, it's a pretty huge company.

With all these big numbers, you can imagine the exponential amount of data Walmart generates. Walmart collects over 2.5 petabytes of data from 1 million customers every hour. Yes, you read that right. Now to make sense of all this information, Walmart has created 'Data Café' – a state-of-the-art analytics hub.

In Data Cafe, over 200 streams of internal and external data, including 40 petabytes of recent transactional data, can be modeled, manipulated, and visualized.

Walmart also constantly analyses over 100 million keywords to know what people near each store are saying on social media. This gives them a better understanding of their customer behavior on what they like and dislike.

Fig.22. walmart data generate

This global chain uses modern tools and technologies to derive business insights and improve customer satisfaction. Some of these technologies include Python, SAS, and NoSQL databases such as Cassandra and Hadoop.

Using all these technologies and data analysis techniques, Walmart can better manage its supply chain, optimize product assortment, personalize the shopping experience, and give relevant product recommendations.

Data analytics for beginners should not merely be theoretical, but also be practical. Data analytics is a lot more practical than theoretical. Hence, here we will have a look at a demo on data analytics for beginners exclusively.

Data analytics technology has become increasingly important in recent years, with more and more organizations relying on it to make informed decisions. However, there are several challenges that must be addressed for data analytics technology to continue to advance and thrive.

One of the main challenges facing data analytics technology is the sheer volume of data being generated. As more devices become connected to the internet and more data is collected, it becomes increasingly difficult to store, process, and analyze this data. This is particularly true in industries like healthcare and finance, where privacy and security concerns must also be taken into account.

Another challenge is the need for skilled professionals who can effectively analyze and interpret data. There is currently a shortage of data scientists and analysts who have the skills and knowledge needed to make sense of complex data sets. This shortage is expected to continue in the coming years, making it more difficult for organizations to leverage the power of data analytics.

In addition to these challenges, there are also concerns about the ethical implications of data analytics. As more data is collected and analyzed, there is a risk that privacy and security could be compromised. There are also concerns about bias in data analytics, as algorithms and models can be influenced by the data they are trained on.
Despite these challenges, the potential for data analytics technology is significant. As more data becomes available, it has the potential to drive innovation, improve decision-making, and enhance efficiency. The use of artificial intelligence and machine learning algorithms can help to automate the data analytics process, making it more efficient and effective.

In the future, we can expect to see data analytics technology continue to evolve and improve. As new technologies emerge and more data becomes available, the potential for data analytics to transform industries and drive innovation will only continue to grow. However, it will be important to address the challenges and concerns surrounding data analytics in order to ensure that it is used in a responsible and ethical manner.

There are several potential futures for data analytics technology that could have a significant impact on society and the economy. One possible future is the continued democratization of data analytics. As more tools and platforms become available, it is becoming easier for non-experts to analyze and interpret data. This could lead to a more data-driven culture, with individuals and organizations making decisions based on data insights rather than intuition or guesswork.

Another potential future is the increasing use of predictive analytics. With the help of machine learning algorithms and artificial intelligence, data analytics could be used to make predictions about future events or trends. This could be particularly useful in industries like finance, where predicting market trends could help investors make more informed decisions.

In addition, we could see the development of more sophisticated data analytics tools that are better able to handle complex data sets. This could involve the use of natural language processing and other advanced technologies to analyze unstructured data like text and images.

However, there are also potential risks associated with the future of data analytics technology. As more data is collected and analyzed, there is a risk that personal privacy could be compromised. There are also concerns about the potential for algorithms and models to perpetuate bias, which could have negative consequences for individuals and society as a whole.

To mitigate these risks, it will be important for organizations and governments to establish clear regulations and ethical guidelines for the use of data analytics. In addition, there will be a need for ongoing education and training to ensure that professionals have the skills and knowledge needed to use data analytics in a responsible and ethical manner.

Overall, the future of data analytics technology is promising, with the potential to transform industries and drive innovation. However, it will be important to address the challenges and risks associated with this technology in order to realize its full potential.

5. Discussion and Implications

The discussion focuses on the implications of the recent advancements in data analytics technologies. The advancements have enabled businesses to gain insights into their data, which has resulted in better decision-making capabilities. The technologies have also enabled businesses to personalize their offerings, optimize their operations, and increase their efficiency. The use of data analytics technologies has also raised concerns around data privacy and security, and it is crucial for businesses to implement appropriate measures to safeguard their data.

5.1. Results analysis and interpretation

After we cleaned up our data. then analysis is required! our goal will have a significant impact on the type of data analysis we perform. But there are a lot of methods out there. Regression analysis, time-series analysis, and univariate or bivariate analysis are a few of them we may be familiar with. But how we use them is more crucial than the various types. What insights we hope to gain will determine how to proceed. All types of data analysis can be categorized broadly into one of the following four groups, like Descriptive analysis, Diagnostic analysis, Predictive analysis and Prescriptive analysis.

Finally after we completed all of our analysis. We have to Sharing these insights with the general public (or at the very least with the stakeholders in your organization) is the last step of the data analytics process. This involves more than just disclosing the unprocessed results of your work; it also entails interpreting the findings and communicating them in a way that is understandable to a variety of audiences. Since you'll be presenting information to decision-makers quite frequently, it's crucial that the conclusions you draw be entirely transparent and unambiguous. Because of this, reports, dashboards, and interactive visualizations are frequently used by data analysts to back up their conclusions.

5.2. A comparison to previous literature

The analysis of recent literature and industry sources has revealed that there have been significant advances in data analytics technologies in recent years. The developments in machine learning algorithms, natural language processing, and deep learning have enabled businesses to analyze large amounts of data quickly and efficiently. These technologies have been applied in various industries such as healthcare, finance, marketing, and e-commerce, resulting in better decision-making capabilities, improved customer experiences, and optimized processes. However, the use of data analytics technologies has also raised concerns around data privacy and security, and businesses must implement appropriate measures to safeguard their data.

5.3. The study's implications and importance

This study's implications and importance are significant as businesses need to stay competitive in today's data-driven world. The report highlights the various technologies that have been developed, which businesses can utilize to gain insights into their data and make data-driven decisions. The report also emphasizes the importance of implementing appropriate measures to safeguard data privacy and security, which is a growing concern among businesses and consumers alike.

5.4. Recommendations and suggestions for future research

Future research can explore the further applications of data analytics technologies in different industries, such as education, transportation, and energy. The research can also examine the development of new data analytics technologies and their potential impact on businesses. Additionally, research can explore the ethical considerations of using data analytics technologies, such as the impact on privacy, security, and social responsibility. Finally, research can investigate the impact of data analytics technologies on job roles and the necessary skills needed for individuals to thrive in a data-driven world.

6. Conclusion

Advances in data analytics technologies have revolutionized the way we process, analyze, and interpret large volumes of data. The emergence of new tools and techniques has made it possible to extract valuable insights from data that was previously considered too complex or voluminous to analyze. This has led to improvements in decision-making, increased efficiency, and enhanced business performance.

The use of machine learning algorithms, artificial intelligence, and natural language processing has enabled us to process and analyze large volumes of data quickly and accurately. This has resulted in the development of predictive models that allow us to forecast future trends and identify potential risks before they become significant problems. Moreover, advances in data visualization techniques have made it possible to present complex data in a simple and understandable format. This has made it easier for decision-makers to identify patterns, trends, and outliers and take appropriate action.

Overall, the advances in data analytics technologies have played a critical role in transforming the way we use data to drive business decisions. As we move forward, it is likely that we will continue to see further developments in this field, leading to even more powerful tools and techniques for processing, analyzing, and interpreting data.

7. References

[1] Lyman P, Varian H. How much information 2003? Tech. Rep, 2004. [Online]. Available: http://www2.sims.berkeley. edu/research/projects/how-much-info-2003/printable_report.pdf.

[2] Xu R, Wunsch D. Clustering. Hoboken: Wiley-IEEE Press; 2009.

[3] Ding C, He X. K-means clustering via principal component analysis. In: Proceedings of the Twenty-first International Conference on Machine Learning, 2004, pp 1–9.

[4] Kollios G, Gunopulos D, Koudas N, Berchtold S. Efficient biased sampling for approximate clustering and outlier detection in large data sets. IEEE Trans Knowl Data Eng. 2003;15(5):1170–87.

[5] Han J. Data mining: concepts and techniques. San Francisco: Morgan Kaufmann Publishers Inc.; 2005.

[6] Neirotti, P., Pesce, D., and Battaglia. 2021. Algorithms for operational decision-making: An absorptive capacity perspective on the process of converting data into relevant knowledge. *Technological Forecasting & Social Change*, 173: 121088.

[7] Tseng, H. T., Aghjaali, N., Hajli, N. (2022). Customer agility and big data analytics in new product context. *Technological Forecasting & Social Change*, 180: 121690.

[8] Sharma, M., sehrawat, R., Daim, T., Shaygan, A. (2021). Technology assessment: Enabling blockchain in hospitality and tourism sectors. *Technological Forecasting & social Change*, 169: 120810.

[9] Wang, J., Xu, C., Zhang, J. and Zhong, R., 2021. Big data analytics for intelligent manufacturing systems: A review. *Journal of Manufacturing Systems*.

[10] Zytek, A., Liu, D., Vaithianathan, R. and Veeramachaneni, K., 2021. Sibyl: Understanding and Addressing the Usability Challenges of Machine Learning In High-Stakes Decision Making. *IEEE Transactions on Visualization and Computer Graphics*, *28*(1), pp.1161-1171.

[11] Phan, T. N. T., Bertrand, P., Vo, X. V., & Jones, K. (2021). Investigating financial decision-making when facing skewed distributions of return: A survey study in Vietnam. *The Quarterly Review of Economics and Finance*.

[12] Ilyina, L. A., Tikhonov, V. S., Sakhno, Y. S., & Viktorova, N. V. (2019). The Mechanism of Highly Effective Decision Making in Modern Business Systems. In *The Leading Practice of Decision Making in Modern Business Systems*. Emerald Publishing Limited.

[13] Caines, J., Bridglall, B. L., & Chatterji, M. (2014). Understanding validity and fairness issues in high-stakes individual testing situations. Quality Assurance in Education.

[14] Gati, I., & Kulcsár, V. (2021). Making better career decisions: From challenges to opportunities. Journal of Vocational Behavior, 126, 103545.

[15] https://www.gartner.com/en/topics/data-and-analytics

[16] https://www.coursera.org/articles/data-analytics

[17] https://www.makeuseof.com/tag/what-is-data-analysis/

[17] Chen, H., Chiang, R. H. L., & Storey, V. C. (2012). Business intelligence and analytics: From big data to big impact. MIS Quarterly, 36(4), 1165-1188.

[18] Davenport, T. H., & Harris, J. G. (2007). Competing on analytics: The new science of winning. Harvard Business Press.

[19] Holsapple, C. W., & Wu, J. (2011). Business intelligence: An introduction to the concepts and IT-based techniques of data analysis. Wiley-Blackwell.

[20] Marr, B. (2019). Artificial intelligence in practice: How 50 successful companies used AI and machine learning to solve problems. John Wiley & Sons.

[21] Wu, X., Zhu, X., Wu, G. Q., & Ding, W. (2014). Data mining with big data. IEEE Transactions on Knowledge and Data Engineering, 26(1), 97-107.

[22] Agneeswaran, V., 2012. Big-data - Theoretical, engineering and analytics perspective. In: Srinivasa, S., Bhatnagar, V. (Eds.), Big Data Analytics. Vol. 7678 of Lecture Notes in Computer Science. Springer Berlin Heidelberg, pp. 8-15.

[23] Agrawal, R., Srikant, R., 1994. Fast algorithms for mining association rules in large databases. In: Proceedings of the 20th International Conference on Very Large Data Bases. (VLDB „94). Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, pp. 487-499.

[24] Anastasia, February 2015. Big data and new product development. Entrepreneurial Insights http://www.entrepreneurial-insights.com/big-data-new-product-development/, Accessed on June 15, 2015.

[25] Assunção, M. D., Calheiros, R. N., Bianchi, S., Netto, M.A.S. , Buyya, R., 2015. Big Data computing and clouds: Trends and future directions, Journal of Parallel and Distributed Computing, Volumes 79–80, May, Pages 3-15.

[26] Bhadani, A., and Chaudhary, S., 2010. Performance evaluation of web servers using central load balancing policy over virtual machines on cloud, Proceedings of the Third Annual ACM Conference, Bangalore, ACM.

[27] Blount, M., Ebling, M., Eklund, J., James, A., McGregor, C., Percival, N., Smith, K., Sow, D., 2010. Real-time analysis for intensive care: Development and deployment of the Artemis analytic system. IEEE Engineering in Medicine and Biology Magazine 29 (2), 110-118.

[28] A. McAfee and E. Brynjolfsson. "Big Data: the management revolution." Harvard business review, Vol. 90, No. 10, pp. 60-68, 2012.

[29] S. Cha and M. Wachowicz. Developing a real-time data analytics framework using Hadoop. 2015 IEEE International Congress on Big Data, pages 657–660, June 2015.

[30] M. Berthold, D. Hand, Intelligent Data Analysis. Springer, Berlin, 2009.

[31] Kordon, A. K. (2020). Applying Data Science: How to Create Value with Artificial Intelligence. Cham, Switzerland: Springer International Publishing. https://doi.org/10.1007/978-3-030-36375-8.

[32] Lu, Huang, Ting-tin Hu, and Hai-shan Chen. "Research on Hadoop Cloud Computing Model and its Applications.". Hangzhou, China: 2012, pp. 59–63, 21-24 Oct. 2012.

[33] Castelino, C., Gandhi, D., Narula, H. G., & Chokshi, N. H. (2014). Integration of Big Data and Cloud Computing. International Journal of Engineering Trends and Technology (IJETT), 100-102.

[34] Pavlopoulos, Georgios A.; Iacucci, Ernesto; Iliopoulos, Ioannis; Bagos, Pantelis (2013). Interpreting the Omics 'era' Data. Multimedia Services in Intelligent Environments. Smart Innovation, Systems and Technologies.

[35] Benjamin B. Bederson and Ben Shneiderman (2003). The Craft of Information Visualization: Readings and Reflections, Morgan Kaufmann ISBN 1-55860-915-6.