

International Journal of Computer and Communication Technology

Volume 9
Issue 1 *Research on Computing and
Communication Sciences.*

Article 3

July 2023

A CREATIVE JOURNEY INTO BIG DATA ANALYSTS

Ravi M

BISHOP HEBER COLLEGE, TIRUCHIRAPALLI, TAMIL NADU, INDIA, cs215114663@bhc.edu.in

Follow this and additional works at: <https://www.interscience.in/ijcct>



Part of the [University Extension Commons](#)

Recommended Citation

M, Ravi (2023) "A CREATIVE JOURNEY INTO BIG DATA ANALYSTS," *International Journal of Computer and Communication Technology*. Vol. 9: Iss. 1, Article 3.

DOI: 10.47893/IJCCT.2023.1435

Available at: <https://www.interscience.in/ijcct/vol9/iss1/3>

This Article is brought to you for free and open access by the Interscience Journals at Interscience Research Network. It has been accepted for inclusion in International Journal of Computer and Communication Technology by an authorized editor of Interscience Research Network. For more information, please contact sritampatnaik@gmail.com.

A CREATIVE JOURNEY INTO BIG DATA ANALYSTS

Cover Page Footnote

ACKNOWLEDGEMENTS: Data sources and providers: Big data analysts may need to acknowledge the organizations, companies, or individuals who provided the data they analyzed. Supervisors and managers: Big data analysts may need to acknowledge their supervisors and managers who provided guidance, resources, and support during the analysis process. Colleagues and team members: Big data analysts may need to acknowledge their colleagues and team members who contributed to the analysis, whether through collaboration or providing feedback. Software and technology providers: Big data analysts may need to acknowledge the software and technology providers who provided tools and resources for data analysis. Academic resources: Big data analysts may need to acknowledge academic resources, such as research papers or textbooks, that informed their analysis methods or theoretical frameworks. Funding agencies: Big data analysts may need to acknowledge the funding agencies or organizations that supported their work, whether through grants, scholarships, or other forms of financial support.

A CREATIVE JOURNEY INTO BIG DATA ANALYSTS

M RAVI , Dr B GAYATHIRI ,B.Sc. cs associate professor ,BHC.

II nd B.Sc. computer science, Bharathidasan university ,

Bishop Heber College ,Trichy ,Tamil Nadu ,India.

cs215114663@bhc.edu.in

gayathri.cs@bhc.edu.in

ABSTRACT:

A big data analyst is a professional who specializes in analyzing and interpreting large and complex sets of data. They use various tools and techniques to extract insights and trends from the data, which can help businesses make informed decisions and improve their operations. To become a big data analyst, you typically need a strong background in statistics, computer science, and data analysis. You may also need experience working with big data tools such as Hadoop, Spark, and NoSQL databases. Additionally, strong communication and collaboration skills are important, as big data analysts often work with teams of other data professionals and stakeholders within the organization. This abstract explores the concept of big data and its significance in today's data-driven world. It delves into the three primary characteristics of big data—volume, velocity, and variety—and highlights the challenges and opportunities they present.

It also emphasizes the importance of effective data management, storage, and processing techniques, such as distributed computing and parallel processing, in handling big data. Furthermore, the abstract discusses various applications of big data across diverse sectors, including healthcare, finance, marketing, and transportation. It outlines how big data analytics can enhance operational efficiency, enable predictive modeling, support personalized recommendations, and facilitate data-driven decision-making. Additionally, it touches upon the ethical considerations and privacy concerns associated with big data, emphasizing the need for responsible data governance and privacy protection.

KEYWORDS:

Data analysis ,Data mining, Machine learning, Statistical analysis, Data visualization, Business intelligence, Predictive modeling, Cloud computing, Artificial intelligence.

INTRODUCTION:

A big data analyst is a professional who works with large and complex sets of data to derive meaningful insights and trends. With the rise of digital technologies, businesses and organizations are collecting vast amounts of data every day, and big data analysts are responsible for analyzing this data to provide insights that can help drive business decisions and improve performance. Big data analysts use various tools and techniques to manage and analyze large data sets, including statistical and machine learning algorithms, data visualization tools, and big data processing frameworks like Hadoop and Spark. The role of a big data analyst requires a strong foundation in statistical analysis, computer science, and data management. They must be able to collect, clean, and process data, as well as analyze and interpret it to identify trends, patterns, and relationships. Big data analysts also need excellent communication and collaboration skills, as they often work with teams of other data professionals and stakeholders within their organization.

LITERATURE OF REVIEW:

[1] "Big Data Analytics in Healthcare: A Literature Review, Classification and Future Trends" by Shahriar Akter, Tasadduq Imam, and Yanchang Zhao (International Journal of Big Data Intelligence, 2015)

[2] "Big Data Analytics: A Literature Review Paper" by Ahmed Alghamdi, Naif Alshahrani, and Abdullah Alghamdi (International Journal of Advanced Computer Science and Applications, 2018)

[3] "Big Data Analytics in Supply Chain Management: A Comprehensive Literature Review and Research Agenda" by Md. Saifuddin Khalid, Xiaohong Liu, and Qingying Li (International Journal of Production Research, 2018)

[4] "Big Data Analytics in E-commerce: A Literature Review and Research Agenda" by Fangfang Zhang, Jun Li, and Xueqin Zhu (International Journal of Information Management, 2018)

[5] "Big Data Analytics for Financial Market Prediction: A Literature Review and Classification Framework" by Tanvi Purohit and Amit Kumar Jain (Journal of Big Data, 2018)

BIG DATA ANALYSTS :

Big data analysts are professionals who specialize in analyzing and

interpreting large and complex data sets using various tools and techniques. They are responsible for collecting, processing, cleaning, and analyzing massive amounts of data to identify patterns, trends, and insights that can help businesses make data-driven decision.

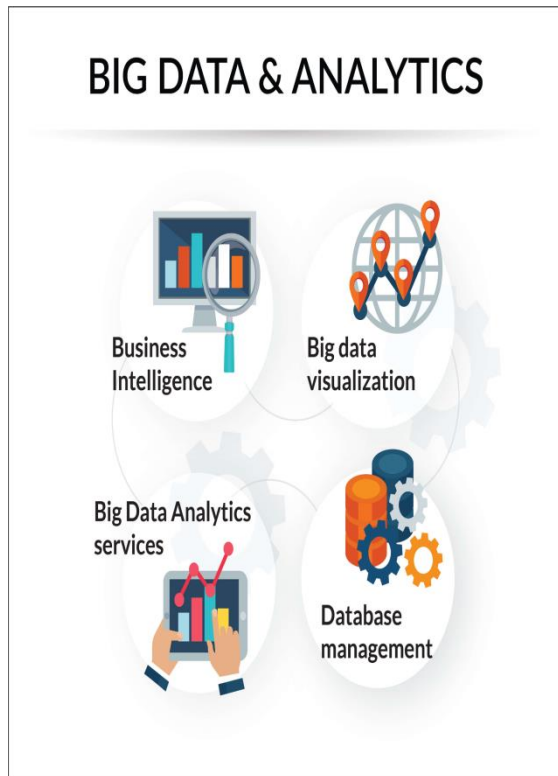


FIG.1 BIG DATA ANALYSTS

Some of the key responsibilities of a big data analyst may include:

Designing and implementing data analysis processes and tools.

Designing and implementing data analysis processes and tools can be a complex task that requires a deep understanding of both the data and the analytical tools and techniques that will be used.

Define the problem: The first step in designing any data analysis process is to clearly define the problem you are trying to solve. What are the questions you want to answer? What are the goals you want to achieve? What are the data sources you will use? Once you have a clear understanding of the problem you are trying to solve, you can move on to the next steps.

Collect and clean the data: Data analysis requires high-quality data. You will need to collect the data you need and clean it up, so it is consistent and accurate. You may need to use data cleaning techniques to remove duplicates, correct errors, or fill in missing data.

Choose the right analytical tools: There are many analytical tools available, and each has its strengths and weaknesses. You will need to choose the right tools for your specific problem. Common tools include statistical analysis tools, machine learning algorithms, and data visualization tools.

Create a data analysis plan: Once you have the data and the tools, you can create a data analysis plan. This plan should outline the steps you will take to analyze the data, the tools you will use, and the expected outcomes.

Implement the plan: With the plan in place, you can start implementing the data analysis process. This will involve using the tools and techniques you have chosen to analyze the data and draw conclusions.

Evaluate and refine the process: Once the analysis is complete, you will need to evaluate the process and refine it as necessary. This may involve revisiting the problem definition, adjusting the analytical tools, or collecting additional data.

Collecting and processing large data sets from various sources.

Collecting and processing large datasets from various sources can be a challenging task, but there are several steps that can be taken to make the process smoother:

Define your research question or objective: It is important to know what you are looking for before you start collecting and processing data. This will help you to determine the type of data you need to collect and the sources you need to tap into.

Identify data sources: Determine the sources of data that are relevant to your research question. These sources can include existing databases, public data sources, or data that needs to be collected manually.

Collect data: Once you have identified the data sources, you need to collect the data. Depending on the source, you may need to use web scraping tools or APIs to extract data.

Clean and preprocess data: After collecting the data, you need to clean and preprocess it. This includes removing any irrelevant data, handling missing data, and converting the data into a format that is suitable for analysis.

Store and manage data: Once the data has been cleaned and preprocessed, it needs to be stored and managed. This can be done using a variety of tools and technologies, including databases, data warehouses, and data lakes.

Analyze and visualize data: The final step in the process is to analyze and visualize the data. This can involve using statistical tools, machine learning algorithms, and visualization techniques to uncover patterns and insights in the data.

Cleaning and transforming data to prepare it for analysis.

Cleaning and transforming data are essential steps in preparing it for analysis.

Identify missing or incomplete data: Identify any missing or incomplete data in

your dataset. Missing data can be problematic because it can affect the results of your analysis. You can deal with missing data in a number of ways, such as deleting the rows with missing data, imputing missing values, or using advanced techniques such as machine learning algorithms.

Identify and deal with outliers: Identify any outliers in your data. Outliers can have a significant impact on the results of your analysis. You can deal with outliers by removing them, transforming them, or using robust statistical methods that are less sensitive to outliers.

Check for inconsistencies and errors: Check your data for inconsistencies and errors. This can involve checking for duplicate data, verifying that the data is in the correct format, and ensuring that any codes or labels are accurate and consistent.

Normalize and standardize the data: Normalize and standardize the data to make it easier to compare across different variables. Normalizing involves scaling the data so that it has a mean of zero and a standard deviation of one. Standardizing involves scaling the data so that it falls within a specific range, such as 0 to 1.

Transform the data: Transform the data as needed to make it suitable for analysis. This can involve transforming categorical data into numerical data, reducing the dimensionality of the data, or applying advanced techniques such as principal component analysis.

Check for multicollinearity: Check your data for multicollinearity, which occurs when two or more independent variables in a regression model are highly correlated. This can be problematic because it can lead to unreliable estimates of the coefficients and inflated standard errors. You can deal with multicollinearity by removing one of the highly correlated variables or by using advanced techniques such as ridge regression or principal component regression.

Analyzing data using statistical and data mining techniques.

Analyzing data using statistical and data mining techniques involves the use of various tools and techniques to extract insights and patterns from large datasets. Here are some common techniques used in statistical and data mining analysis:

Descriptive statistics: This involves the use of measures such as mean, median,

mode, standard deviation, and variance to summarize and describe the data.

Inferential statistics: This involves making predictions or drawing conclusions about a population based on a sample of data. This can involve hypothesis testing, confidence intervals, and regression analysis.

Data visualization: This involves the use of charts, graphs, and other visual representations to help identify patterns and trends in the data.

Clustering: This involves grouping data points based on similarities and differences in their characteristics. This can help identify subgroups within the data.

Classification: This involves building models that can predict the class of a new data point based on its characteristics. This can be useful in identifying patterns and trends in the data.

Association rule mining: This involves identifying relationships between variables in the data. This can help identify patterns and trends that may not be immediately apparent.

Time series analysis: This involves analyzing data over time to identify

patterns and trends. This can be useful in forecasting future trends.

Identifying patterns and trends in data

Identifying patterns and trends in data is an important task in data analysis and can provide valuable insights for decision-making. There are various techniques and tools that can be used to identify patterns and trends in data, including:

Data visualization: This involves representing data in visual formats such as charts, graphs, and maps, which can help identify patterns and trends.

Statistical analysis: Statistical techniques such as regression analysis, time series analysis, and correlation analysis can be used to identify patterns and trends in data.

Machine learning: Machine learning algorithms can be used to identify patterns and trends in data, such as clustering algorithms, decision trees, and neural networks.

Data mining: Data mining techniques can be used to discover hidden patterns and relationships in large datasets, such as association rule mining and anomaly detection.

Creating data visualizations and reports to communicate findings to stakeholders

Creating data visualizations and reports to communicate findings to stakeholders is an essential part of the data analysis process. The purpose of data visualizations is to convey complex information in a simple and intuitive manner. When designing data visualizations and reports, it is important to keep in mind the audience and the purpose of the communication. Here are some steps that can help you create effective data visualizations and reports:

Define the problem: Start by defining the problem you are trying to solve and the questions you want to answer. This will help you determine what data to collect and analyze.

Gather and clean the data: Collect the data you need and clean it to ensure it is accurate and complete. You can use tools like Excel or Python to clean and manipulate the data.

Choose the right visualization: Select the appropriate visualization based on the type of data and the questions you are trying to answer. For example, if you are looking at trends over time, a line chart may be the best option.

Design the visualization: Once you have chosen the right visualization, design it in a way that is easy to understand and visually appealing. This may involve choosing the right colors, fonts, and layout.

Interpret the data: Interpret the data and provide insights that are relevant to the problem you are trying to solve. Use simple and concise language to explain your findings.

Communicate the results: Finally, communicate your findings to stakeholders in a way that is easy to understand and relevant to their needs. This may involve creating a report or a presentation that summarizes your findings and provides actionable recommendations.

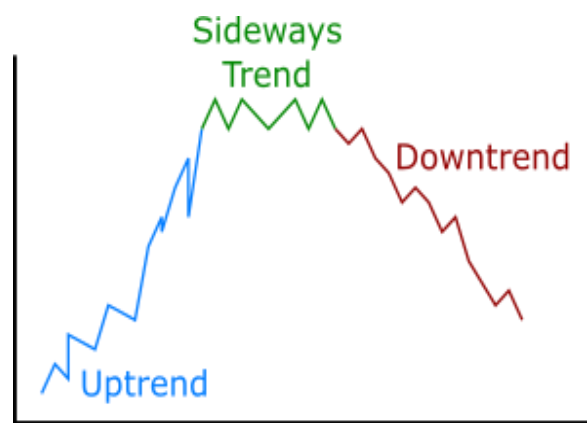


FIG.3 PATTERN AND TRENDS IN DATA

Providing insights and recommendations to support business decisions

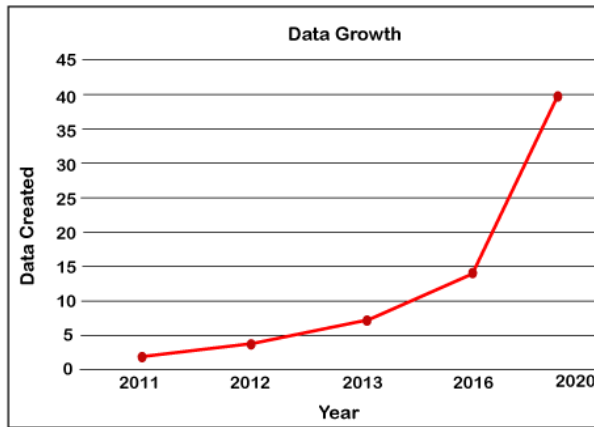


FIG.2 DATA GROWTH

Market Research: Conducting market research is crucial to understand the market trends, customer behavior, and competitors' strategies. You can use my natural language processing (NLP) capabilities to analyze customer reviews, social media posts, and other sources of data to gather insights about what your customers want and what they don't like about your products or services. This can help you make informed decisions about product development, pricing, and marketing strategies.

Financial Analysis: You can use my machine learning algorithms to analyze your financial data and provide insights on your revenue, expenses, cash flow, and profitability. I can help you identify the areas where you are spending too much money, where you can reduce costs, and where you should invest more.

Supply Chain Management: I can help you optimize your supply chain by analyzing data from various sources, such as inventory levels, shipping times, and production schedules. This can help you identify bottlenecks in your supply chain and find ways to reduce lead times, improve quality, and lower costs.

Human Resources: I can help you analyze your HR data to identify patterns and trends in employee turnover, productivity, and engagement. This can help you make informed decisions about training, hiring, and retention strategies.

Marketing and Advertising: You can use my predictive analytics capabilities to analyze customer behavior and predict which marketing campaigns will be most effective.

To become a big data analyst, you typically need a strong foundation in statistics, mathematics, and programming, as well as experience working with big data technologies such as Hadoop, Spark, and NoSQL databases. Many big data analysts hold degrees in fields such as computer science, statistics, mathematics, or a related field, and have experience working with data in various industries.

ACKNOWLEDGEMENTS:

Data sources and providers: Big data analysts may need to acknowledge the organizations, companies, or individuals who provided the data they analyzed.

Supervisors and managers: Big data analysts may need to acknowledge their supervisors and managers who provided guidance, resources, and support during the analysis process.

Colleagues and team members: Big data analysts may need to acknowledge their colleagues and team members who contributed to the analysis, whether through collaboration or providing feedback.

Software and technology providers: Big data analysts may need to acknowledge the software and technology providers who provided tools and resources for data analysis.

Academic resources: Big data analysts may need to acknowledge academic resources, such as research papers or textbooks, that informed their analysis methods or theoretical frameworks.

Funding agencies: Big data analysts may need to acknowledge the funding agencies or organizations that supported their work, whether through grants, scholarships, or other forms of financial support.

CONCLUSION:

Big data analytics has become an essential tool for businesses in many industries to make informed decisions based on large volumes of data. By analyzing data from various sources, businesses can gain valuable insights into customer behavior, market trends, and

operational efficiency, which can lead to improved decision-making and better business outcomes. The use of advanced technologies like machine learning, artificial intelligence, and natural language processing can further enhance the accuracy and speed of data analysis, providing even more valuable insights. Ultimately, big data analytics can help businesses stay competitive in a rapidly changing marketplace by identifying opportunities for growth and improvement.

BOOK:

[1]"Data Smart: Using Data Science to Transform Information into Insight" by John W. Foreman

[2]"Big Data: A Revolution That Will Transform How We Live, Work, and Think" by Viktor Mayer-Schönberger and Kenneth Cukier

[3]"Data Analytics Made Accessible" by Anil Maheshwari

[4]"Data Science for Business: What You Need to Know about Data Mining and

Data-Analytic Thinking" by Foster Provost and Tom Fawcett

[5] "Big Data Analytics: From Strategic Planning to Enterprise Integration with Tools,

Techniques, NoSQL, and Graph" by David Loshin

[6]"Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython" by Wes McKinney

[7]"Data Mining: Practical Machine Learning Tools and Techniques" by Ian H. Witten, Eibe Frank, and Mark A. Hall.

[8]"The Elements of Statistical Learning: Data Mining, Inference, and Prediction" by Trevor Hastie, Robert Tibshirani, and Jerome Friedman.