International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 04 | April - 2023 Impact Factor: 8.176 ISSN: 2582-3930

Cloud-Enabled Data Analytics: A Critical Review of Trends and Technologies

Raghav Saboo[1], Prakhar Gupta[2], Pushpendra Singh[3] ,Assist. Prof. Varsha Kothari[4]

[1] Student, Dept. of Computer Science Engineering, Medi-Caps University, Madhya Pradesh, India

[2] Student, Dept. of Computer Science Engineering, Medi-Caps University, Madhya Pradesh, India.

[3] Student, Dept. of Computer Science Engineering, Medi-Caps University, Madhya Pradesh, India.

[4]Assistant Professor, Dept. of Computer Science Engineering, Medi-Caps University, Madhya Pradesh, India.

Abstract -*Cloud-enabled data analytics is the process of analyzing large data sets using cloud computing resources. It has become increasingly popular due to the vast amount of data that businesses generate and collect. This review paper aims to provide an overview of the current state of cloud-enabled data analytics and its applications.The paper begins with a discussion of the basics of cloud computing and its role in enabling data analytics. It then covers the various types of cloud computing models, including Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Software as a Service (SaaS), and how each model can be used for data analytics.Next, the paper explores the different types of data analytics techniques that can be used in the cloud, including descriptive, predictive, and prescriptive analytics. It also covers the tools and technologies that are commonly used in cloud-based data analytics, such as Apache Hadoop, Spark, and NoSQL databases.The paper then discusses the benefits of cloud-enabled data analytics, including cost savings, scalability, and flexibility. It also covers some of the challenges associated with cloud-enabled data analytics, such as data privacy and security concerns.Finally, the paper provides an overview of some of the applications of cloud-enabled data analytics in various industries, including healthcare, finance, and retail. It also discusses some of the emerging trends in cloud-enabled data analytics, such as edge computing and machine learning.*

Key Words:Cloud computing, Data analytics, Big data, Artificial intelligence, Distributed computing

1. INTRODUCTION

The volume and complexity of data being generated by businesses, organizations, and individuals are increasing rapidly. The amount of data being produced globally is expected to reach 175 zettabytes by 2025, which is an increase from 33 zettabytes in 2018 (Seagate, 2018). This explosive growth of data presents numerous challenges and opportunities for businesses. Data analytics has emerged as a key solution to manage, process, and analyze this vast amount of data, providing valuable insights that can help drive business growth, efficiency, and competitiveness.

Cloud-enabled data analytics has become increasingly popular in recent years, as businesses look to leverage the benefits of cloud computing to process and analyze their data. Cloud-enabled data analytics is the process of analyzing large data sets using cloud computing resources. It offers numerous advantages over traditional on-premise data analytics solutions, including cost savings, scalability, flexibility, and accessibility.

This review paper provides an overview of cloud-enabled data analytics, its applications, benefits, and challenges. It aims to provide a comprehensive analysis of the current state of cloud-enabled data analytics, examine emerging trends, and identify key challenges and opportunities.

Why this Paper is Worth Reading ?

This review paper is worth reading for business leaders, data analysts, researchers, and students who want to gain a deeper understanding of cloud-enabled data analytics. It provides an overview of the key concepts, tools, and technologies used in cloud-enabled data analytics and explores the benefits and challenges associated with this approach.

By reading this paper, you will gain insights into the different types of cloud computing models used for data analytics, including Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Software as a Service (SaaS). You will

also learn about the different types of data analytics techniques used in the cloud, including descriptive, predictive, and prescriptive analytics.

The paper covers the various tools and technologies used in cloud-enabled data analytics, including Apache Hadoop, Spark, and NoSQL databases. It also discusses the benefits of cloud-enabled data analytics, including cost savings, scalability, and flexibility.

Moreover, the paper examines the potential challenges and risks associated with cloud-enabled data analytics, such as data privacy and security concerns. It provides recommendations on how businesses can address these challenges and mitigate the risks.

The paper also explores the applications of cloud-enabled data analytics in various industries, such as healthcare, finance, and retail. It provides examples of how businesses are using cloud-enabled data analytics to gain insights into their customers, optimize their operations, and improve their decision-making processes.

Finally, the paper identifies emerging trends in cloud-enabled data analytics, such as edge computing and machine learning, and discusses their potential implications for businesses.

We also explored the applications of cloud-enabled data analytics in various industries, including healthcare, finance, and retail. We provided examples of how businesses are using cloud-enabled data analytics to gain insights

2. LITERATURE REVIEW

2.1 Data Analytics

Analysis of data is a process of inspecting, cleaning, transforming, and modelling data with the goal of highlighting useful information, suggesting conclusions, and supporting decision making. Data analysis has multiple facets and approaches, encompassing diverse techniques under a variety of names, in different business, science, and social science domains.

Data mining is a particular data analysis technique that focuses on modelling and knowledge discovery for predictive rather than purely descriptive purposes. Business intelligence covers data analysis that relies heavily on aggregation, focusing on business information. In statistical applications, some people divide data analysis into descriptive statistics, exploratory data analysis (EDA), and confirmatory data analysis (CDA). EDA focuses on discovering new features in the data and CDA on confirming or falsifying existing hypotheses. Predictive analytics focuses on application of statistical or structural models for predictive forecasting or classification, while text analytics applies statistical, linguistic, and structural techniques to extract and classify information from textual sources, a species of unstructured data. All are varieties of data analysis. Data integration is a precursor to data analysis, and data analysis is closely linked to data visualization and data dissemination.

The available data analysis tools are mostly a collection of data analysis methods that require experts as users. The users need domain knowledge and also need to know which data analysis methods have to be applied to a given problem and which technique meets the requirements for the solution. The expert should also know how the data has to be prepared for the chosen technique and finally, how the technique needs to be configured.

Business users require a much more user- or problem oriented approach to data analysis. Rather than knowing analysis methods, they are experts in the data domain and they know what they want to achieve with data analysis. If they only knew how. They might know, for example, that they want to classify insurance claims as fraudulent or non-fraudulent, given historic information of the customer and the current case. They might want to understand, how the analysis method actually classifies customers (e.g. with a rule set), they might require a certain classification accuracy and that the algorithm is so simple that it can be implemented as an SQL query. Ideally, such users would simply like to feed all these high-level requirements and the data into a tool that would then automatically find the best algorithm in terms of requirements, configure it, run it and create a software module that can be plugged into the business application.

In this paper, we focus on a way to select the most appropriate data analysis algorithm given a problem definition, a set of requirements and a data file.

2.1.1 NEED OF DATA ANALYSIS MODELS

As companies adopt analytics as the new science of winning, organizations will need to focus both on the creation and consumption of insights to enable better decisions. There is need of data analysis due the following reasons:

• The business problem is not clear: In a rush to jump on the analytics bandwagon, business practitioners often forget that the business problem needs to be well- defined for the analytics solution to be relevant to the problem at hand.

• Appropriate stakeholder(s) are not involved: If a firm is using analytics to design a promotion campaign for a certain product, the demand planning teams need to know what's changing to get the product on the shelves. Like any project

team, the right stakeholders need to be involved at the right time. This is especially true when multiple functional groups are involved in a specific business problem.

• Mystery math: With the explosion in data and the availability of technologies that bring applied math to the analytics workbench, analytics practitioners begin to regard the technical analysis as an end in itself. Mathematical techniques are tools necessary to solve the business problem at hand.

• The right expectations are not set: Sophisticated mathematical techniques are often expected to act as magic wands, solving any and every problem at hand. More often than not, this creates unreasonable expectations. As the key sponsor of a failed forecasting project famously said, "Why should there be any error in the forecast if you have used sophisticated mathematical techniques?" This was clearly a case of a mismatch in expectations – it was never communicated to the executive that no mathematical technique, however sophisticated, could accurately predict the future.

• Lack of continuity: As basic a management principle as it may sound, the best analytics ideas tend to lose advantage and diminish in value, due to a variety of reasons ranging from internal organization changes to getting lost in the shuffle of organizational initiatives.

The creation of insights requires a holistic perspective of Descriptive Analytics, Inquisitive Analytics, Predictive Analytics and Prescriptive Analytics:

1. **Descriptive analytics** answers the questions "What happened in the business?" It is looking at data and information to describe the current business situation in a way that trends, patterns and exceptions become apparent
2. **Inquisitive analytics** answers the question "Why is something happening in the business?" It is the study of data to validate/reject business hypotheses
3. **Predictive analytics** answers the question "What is likely to happen in the future". It is data modeling to determine future possibilities
4. **Prescriptive analytics** is the combination of the above to provide answers to the "so what?" and the "now what?" questions. For example, what should I do to retain my key customers? How do I improve my supply chain to enhance service levels while reducing my costs?

The type of analysis problem restricts the list of applicable data analysis techniques to that problem. By the term analysis problem, we mean, whether it is a classification problem, function approximation like time series prediction, a clustering problem, if it is about finding dependencies or associations etc. The second category of requirements is concerned with preferences regarding the solution. These comprise properties like accuracy and simplicity of the solution, if the method is adaptable to new, whether it offers an explanation facility like rule- based systems or functional models like linear regression and how simple the explanation should be. Finally, the data might constrain the applicability of methods. The number of data records, for example, might be too small for some statistical methods, or generally, some methods might cope better with certain types of data than others.

Depending on the type of user, the level at which the requirements of the problem are defined will vary considerably. Some users may understand the difference between function approximation and classification. Thus, there is a need of hierarchical approaches where requirements are iteratively mapped onto lower level requirements until the lowest level is reached.

To choose the analysis methods, the requirements have to be mapped onto properties of the methods and the various stages or steps of data analysis have to be followed.

2.1.2 TECHNIQUES OF DATA ANALYSIS

Conjoint Analysis / Choice Modeling:

Definition-Enables the breakdown of consumer preferences for a good or service into tradeoffs among its component features for the context in which overall assessments are made. Conjoint analysis, a well-liked method for assessing multi-attribute consumer preferences used in market research, is a useful tool for concurrently assessing many gamut mapping techniques. Conjoint analysis' goal is to ascertain which pairing of a select few traits is most valued by customers. A multi-attribute compositional model called conjoint analysis.

Application - Optimizing product layout, researching pricing elasticities for demand, simulating consumer reactions to new or updated products, and identifying competitive advantages and disadvantages.

Pros - Of all survey research methods, this one most accurately simulates the actual purchasing process. Being able to perform "what if" scenarios that haven't been expressly evaluated gives it flexibility. Excellent for developing new products and setting prices. Conjoint analysis aids in identifying a product or service's ideal attributes.

Cons - Models focus more on preference share than market share.There are restrictions on how many features can be used in a study.

Factor Analysis:

Definition - Finds a set of unobserved structure in the data by identifying a set of underlying dimensions (or "Factors") inside a collection of variables. The phrase "factor analysis" was initially used by Thurstone and is used as a tool for data reduction or structure finding. A multivariate statistical exploratory technique is factor analysis. It is used to condense the data from a big collection of variables into a more manageable collection of composite variables known as FACTORS.

Application - Discovering conceptual or benefit factors underlying expressed product impressions and preferences; reducing the number of variables for study.

Pros - Simplifies extensive or complex sets of variables or attributes. can be utilised to comprehend the customer's thought process. commonly applied to subjective metrics like product attribute assessments, opinions, and beliefs.

Cons - A part of the problem is the results' subjective interpretation. is frequently a supporter of other studies, like segmentation, rather than a goal in and of itself.

Cluster Analysis:

Definition - Cluster analysis is an exploratory data analysis tool for solving classification problems. Its object is to sort cases (people, things, events, etc.) into groups, or clusters, so that the degree of association is strong between members of the same cluster and weak between members of different clusters. A cluster is a group of relatively homogenous cases or observations. Each cluster thus describes, in terms of the data collected, the class to which its members belong; and this description may be abstracted through use from the particular to the general class or type. Uses any of several techniques (viz. Nearest Neighbors, K-Means etc.) to classify people, objects, or variables into more homogeneous groups.

Application - Identifying / describing market segments; developing typological findings and describing target markets.

Pros - Allows a deeper understanding of the market. Can greatly aid messaging and new product development by targeting homogeneous groups.

Cons - Subjective interpretation of the results is a component. The technique is mathematical and therefore has no underlying model against which to test statistical hypotheses. K-means is a fast cluster analysis method, in which accuracy depends on the use of initialization algorithms that are usually serial and slow.

Analysis: Regression

Defined as the study of a single interval scale variable's dependency on one (simple) or more (many) variables, such as market share. In regression analysis, the dependent and independent variables are clearly related, and the measurement error variance is estimated. An essential method for assessing and forecasting data with categorical variables is logistic regression. A crucial statistical technique for categorical data modelling and prediction is logistic regression.

Application - Predicting sales, market share, and profitability; simulating consumer behaviour and the effects of marketing initiatives; and calculating the elasticity and reaction functions.

Pros - Fantastic tool for predictive modelling. a tried-and-true approach. Diagnostics can be used to assess the model's effectiveness.

Cons - Highly correlated and outlier-prone data. Premature convergence and slow convergence speed are the two main issues with regression analysis approaches. Real-world data mining applications frequently present us with the challenge of doing logistic regression analysis without having access to the entire set of data up front. [1]

2.2 Big Data

Big data is the new term that contains large and complex datasets. It is difficult to manage these datasets without new technology. The Mckinsey Global Institute (MGI) published a report on big data that describes the various business opportunities that big data opens. Paulo Boldi, One of the authors says “Big Data does not need big machines, it needs big intelligence”. There are two types of Big Data is as follows:

Structured Data: These data can be easily analyzed. It is in numerical form, figures, and transaction data etc.

Unstructured Data: These data contain complex information such as Email attachments, Images comments on social networking sites. These data cannot be easily analyzed.

Doug Lancy was the first one talking about 3v’s in big data management [2]:

- Volume - It describes the amount of data. It refers to mass quantities of data.
- Variety - It describes different types of data and sources including structured, semi-structured and unstructured data.
- Velocity - It defines the motion of data. Data created rapidly, processed and analyzed.

2.3 Cloud Computing

2.3.1 Concept of Cloud Computing

Distributed computing concerns the provisioning of assets, including calculation, memory, stockpiling, organization, and applications/administrations, over the Internet. This figuring worldview essentially embraces the customer worker engineering and works with unified sending and calculation offloading for applications. Along these lines, distributed computing is cost-productive in application sending and upkeep, and adaptable in asset provisioning and in decoupling administrations from fundamental advancements at both the customer and worker side. Distributed computing and its empowering advancements have been read for quite a long time, and various develop figuring stages have been conveyed on the lookout, e.g., Amazon EC2, Google Cloud Platform, Microsoft Azure and IBM SmartCloud. The speedy development of versatility empowering innovations alongside the ubiquity of cell phones these days has pushed the exploration on distributed computing to help portable applications just as client/gadget portability. Current cell phones are furnished with incredible detecting abilities, henceforth, can give tactile information of their encompassing zones. By abusing such information, the gadgets andapplications can bring setting mindful administrations to clients. Due to this pattern, versatile distributed computing has been presented in as a reconciliation of portable registering and distributed computing. It is officially characterized as a novel registering worldview focusing on asset provisioning to the gadgets to help the setting mindfulness ability of both the gadgets and applications. The work additionally studies various stages that have been created for this figuring worldview.

Cloud computing is generally considered to have the following characteristics:

Virtualization: Virtualization is the core technology of cloud computing, and many other features that depend on it. The application of virtualization technology can integrate heterogeneous computing resources to form a resource pool for users to access.

Service-oriented: Cloud computing provides three levels of services, namely Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Software as a Service (SaaS). IaaS is the lowest-level service that directly provides compute, memory, and networking equipment. Users have the greatest degree of freedom and can build their own platforms and software. PaaS is one level higher than IaaS, providing a ready-made cloud platform, saving the work of developing the platform. SaaS provides more convenient services; users can directly use the provided software without any development.

Elasticity and scalability: The cloud scale can be easily expanded without affecting the cloud services currently provided externally. Resources in the cloud are infinitely desirable to users and can be automatically provisioned and reclaimed quickly on demand.

Reliable and universal: Cloud computing technology provides a variety of fault-tolerant mechanisms to ensure high reliability of services. Data is placed with multiple copies to prevent data loss due to hardware failure. Computer services that were stopped due to hardware failures can still continue elsewhere through virtual machine migration. Virtualization makes cloud computing resources transparent to users and supports applications in different industries at the same time.

Economies of scale: The cloud computing platform does not have high requirements for hardware facilities, and a large number of idle ordinary computers can be integrated into the resource pool through virtualization. For users, it saves hardware costs and daily management costs of self-built platforms. For cloud service providers, the versatility of cloud computing has greatly improved the utilization of resources, and the scale has significantly increased economic benefits [2].

2.5 Machine Learning

AI is the order of instructing PCs to anticipate results or arrange objects without being unequivocally customized for such undertakings. One of its fundamental suspicions is that it is feasible to construct calculations that can foresee future, beforehand inconspicuous qualities utilizing prepared information and the utilization of measurable methods. AI has been profoundly fruitful in zones such as self-driving vehicles, discourse acknowledgment, compelling web search, and buy suggestions, to give some examples models. This achievement is generally because of the accessibility of huge datasets and the persistent upgrades in the computational force of workers and GPUs. AI calculations can be arranged into two principal gatherings: regulated and non-supervised calculations. Regulated learning alludes to building models given an assortment of preparing indicators X1, X2, . . ., Xp and the

relating reaction variable Y, though in solo realizing there exist just indicators, consequently the calculations need to become familiar with the design of the preparation information (bunching). At the point when the objective is to anticipate a nonstop or quantitative yield esteem, the relating issue to be addressed is called relapse, while the expectation of a downright or subjective yield is known as an arrangement issue. In Fig. we give a scientific categorization of the absolute most mainstream AI calculations utilized by and by.

Machine learning methods can be parametric where certain assumptions are made about the functional form of the model and training data is then used to fit its parameters, e.g., as in polynomial regression, or non- parametric, e.g., neural networks. Machine learning can be used also for inference tasks, i.e., in order to understand how the response variable is affected when the predictors change[3].

3. Pros of Cloud-Enabled Data Analytics:

1. **Scalability**: Cloud-enabled data analytics provides scalability, allowing businesses to quickly scale up or down their analytics infrastructure based on their needs. This eliminates the need for businesses to invest in expensive hardware or software upgrades, as the cloud provider can handle the scaling automatically. According to a report by MarketsandMarkets, the cloud analytics market is expected to grow from $13.2 billion in 2018 to $23.2 billion by 2023, at a CAGR of 11.9%[4].

2. **Cost Savings:** Cloud-enabled data analytics can help businesses save money on infrastructure costs, as they do not need to invest in expensive hardware or software. Additionally, businesses only pay for the resources they use, making it more cost-effective. According to a study by TechNavio, the adoption of cloud analytics solutions can help businesses reduce their IT infrastructure costs by up to 50%[5].

3. **Agility:** Cloud-enabled data analytics provides businesses with the ability to quickly and easily access data from multiple sources and analyze it in real-time. This enables businesses to make data-driven decisions faster and stay competitive. According to a report by Forbes, businesses that adopt cloud analytics are 2.5 times more likely to be able to make decisions in real-time than businesses that do not[6].

4. **Accessibility:** Cloud-enabled data analytics provides businesses with the ability to access data from anywhere, as long as there is an internet connection. This enables businesses to work remotely and collaborate with colleagues from different locations. According to a study by IDG, 69% of businesses that have adopted cloud analytics have reported improved collaboration and teamwork[7].

5. **Security:** Cloud-enabled data analytics providers typically offer high levels of security, ensuring that data is encrypted and protected from unauthorized access. According to a report by Gartner, by 2022, the majority of cloud analytics providers will have enhanced security measures, including advanced threat detection and multi-factor authentication[8].

Cons of Cloud-Enabled Data Analytics:

1. **Dependency on Internet Connectivity: Cloud**-enabled data analytics requires a reliable internet connection, without which businesses may not be able to access their data or run their analytics processes. According to a study by Frost & Sullivan, unreliable internet connectivity is one of the biggest challenges faced by businesses that adopt cloud analytics[9].

2. **Data Security Concerns:** While cloud-enabled data analytics providers offer high levels of security, businesses may still have concerns about the security of their data when it is stored on the cloud. According to a study by PwC, 69% of businesses that adopt cloud analytics are concerned about the security of their data[10].

3. **Data Governance and Compliance:** Businesses must ensure that their data governance and compliance policies are aligned with the cloud-enabled data analytics provider's policies. According to a report by Forrester, businesses that adopt cloud analytics must ensure that they have a clear understanding of their provider's data governance policies, and that these policies comply with their own internal policies and industry regulations[11].

4. **Vendor Lock-In:** Businesses may become locked into a particular cloud-enabled data analytics provider, which could limit their options for future expansion or innovation. According to a report by Gartner, businesses that adopt cloud analytics should ensure that they have the ability to easily migrate their data and processes to another provider, in case they are not satisfied with their current provider[8].

5. **Performance and Latency:** Cloud-enabled data analytics may experience latency issues, especially when dealing with large data sets. Additionally, the performance of cloud-enabled data analytics may be affected by factors outside of the business's control, such as network congestion or hardware failures. According to

a report by TechTarget, businesses that adopt cloud analytics must ensure that their provider has adequate resources to handle their workload, and that they have a plan in place for addressing latency and performance issues[12].

4. Discussion:

Cloud-enabled data analytics is a rapidly evolving field that combines cloud computing infrastructure and services with data analytics techniques to process and analyze large amounts of data. Cloud-enabled data analytics offers several benefits, including scalability, cost-effectiveness, and accessibility.

Scalability is one of the most significant advantages of cloud-enabled data analytics. Cloud infrastructure can quickly and easily scale up or down based on the demand for processing power, storage, and other resources. This means that organizations can quickly and easily increase or decrease their computing resources to accommodate changes in their data analytics needs.

Cost-effectiveness is another advantage of cloud-enabled data analytics. With cloud computing, organizations only pay for the computing resources they use, which makes it more cost-effective than investing in expensive hardware and software infrastructure. Moreover, cloud-enabled data analytics can reduce costs associated with maintenance, upgrades, and licensing fees.

Accessibility is another benefit of cloud-enabled data analytics. Cloud services can be accessed from anywhere, as long as there is an internet connection, which means that organizations can easily collaborate with partners and colleagues across the globe. This accessibility can help organizations to leverage data analytics tools to improve decision-making processes and enhance business operations.

However, there are also challenges associated with cloud-enabled data analytics, such as data security and privacy concerns, data quality issues, and the need for specialized skills and expertise. These challenges can be addressed through effective data governance policies, compliance with data protection regulations, and developing innovative data cleaning and integration techniques.

In conclusion, cloud-enabled data analytics offers several advantages that can help organizations to leverage the benefits of data analytics while minimizing costs and increasing accessibility. However, organizations must also address the challenges associated with this field to ensure that their data analytics processes are effective, efficient, and secure.

REFERENCES

[1]. A Comparative Study of Data Analysis Techniques by Prateek Bihani and S. T. Patil
[2]. A Review Paper on Big Data Analytics by Ankita S. Tiwarkhede , Prof. Vinit Kakde
[3]. A Review of Machine Learning Algorithms for Cloud Computing by Abhishek Dwivedi, Shekhar Verma
[4]. https://www.marketsandmarkets.com/Market-Reports/cloud-based-business-analytics-market-959.html
[5]. https://www.technavio.com/report/cloud-computing-market-size-industry-analysis
[6]. https://www.forbes.com/sites/forbestechcouncil/2020/10/14/five-reasons-more-businesses-are-choosing-cloud/?sh=41d806f233d9
[7]. https://wadic.net/offshore-software-development-emerging-trends-in-2019/
[8]. https://www.gartner.com/smarterwithgartner/6-steps-for-planning-a-cloud-strategy
[9]. https://www.frost.com/frost-perspectives/challenges-of-adopting-edge-computing/

[10].https://www.pwc.com/gx/en/news-room/press-releases/2022/global-digital-trust-insights-survey.html

[11].https://www.forrester.com/staticassets/glossary.ht_ml

[12].https://www.techtarget.com/searchcloudcomputing/definition/cloud-computing